



Modélisation statistique pour la prédiction du pronostic de patients atteints d'un Accident Vasculaire Cérébral

Brice Ozenne

► To cite this version:

Brice Ozenne. Modélisation statistique pour la prédiction du pronostic de patients atteints d'un Accident Vasculaire Cérébral. Bio-Informatique, Biologie Systémique [q-bio.QM]. Université Claude Bernard - Lyon I, 2015. Français. NNT: 2015LYO10169 . tel-01233049

HAL Id: tel-01233049

<https://theses.hal.science/tel-01233049>

Submitted on 24 Nov 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Université Lyon 1
École doctorale Ecosystèmes Evolution Modélisation Microbiologie
pour obtenir le titre de DOCTEUR
spécialité BIOSTATISTIQUE

par
Brice OZENNE

Modélisation statistique pour la prédiction du pronostic de patients atteints d'un Accident Vasculaire Cérébral

Statistical modeling for predicting the prognosis of stroke patients

À soutenir le 23 octobre 2015 devant la Commission d'Examen

JURY

<i>Directeur de thèse</i>	Docteur	Delphine	MAUCORT-BOULCH, Université Lyon 1
<i>Co-directeur de thèse</i>	Professeur	Norbert	NIGHOGHOSSIAN, Université Lyon 1
<i>Rapporteurs</i>	Directeur de recherche	Florence	FORBES, INRIA Grenoble
	Professeur	Denis	VIVIEN, Université de Caen
<i>Examineurs</i>	Docteur	Martyn	PLUMMER, CIRC
	Docteur	Tae Hee	CHO, Université Lyon 1

UNIVERSITÉ CLAUDE BERNARD - LYON 1

Président de l'Université

Vice-président du Conseil d'Administration

Vice-président du Conseil des Etudes et de la Vie
Universitaire

Vice-président du Conseil Scientifique

Directeur Général des Services

M. François-Noël GILLY

M. le Professeur Hamda BEN HADID

M. le Professeur Philippe LALLE

M. le Professeur Germain GILLET

M. Alain HELLEU

COMPOSANTES SANTÉ

Faculté de Médecine Lyon Est – Claude Bernard

Faculté de Médecine et de Maïeutique

Lyon Sud – Charles Mérieux

Faculté d'Odontologie

Institut des Sciences Pharmaceutiques et
Biologiques

Institut des Sciences et Techniques de la
Réadaptation

Département de formation et Centre de Recherche
en Biologie Humaine

Directeur : M. le Professeur J. ETIENNE

Directeur : Mme la Professeure C. BURILLON

Directeur : M. le Professeur D. BOURGEOIS

Directeur : Mme la Professeure C.
VINCIGUERRA

Directeur : M. le Professeur Y. MATILLON

Directeur : Mme. la Professeure A-M. SCHOTT

COMPOSANTES ET DÉPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies

Département Biologie

Département Chimie Biochimie

Département GEP

Département Informatique

Département Mathématiques

Département Mécanique

Département Physique

UFR Sciences et Techniques des Activités
Physiques et Sportives

Observatoire des Sciences de l'Univers de Lyon

Polytech Lyon

Ecole Supérieure de Chimie Physique Electronique

Institut Universitaire de Technologie de Lyon 1

Ecole Supérieure du Professorat et de l'Education

Institut de Science Financière et d'Assurances

Directeur : M. F. DE MARCHI

Directeur : M. le Professeur F. FLEURY

Directeur : Mme Caroline FELIX

Directeur : M. Hassan HAMMOURI

Directeur : M. le Professeur S. AKKOUCHE

Directeur : M. le Professeur Georges TOMANOV

Directeur : M. le Professeur H. BEN HADID

Directeur : M. Jean-Claude PLENET

Directeur : M. Y.VANPOULLE

Directeur : M. B. GUIDERDONI

Directeur : M. P. FOURNIER

Directeur : M. G. PIGNAULT

Directeur : M. le Professeur C. VITON

Directeur : M. le Professeur A. MOUGNIOTTE

Directeur : M. N. LEBOSNE

Résumé

L'Accident Vasculaire Cérébral (AVC) est une maladie grave pour laquelle des critères très stricts encadrent l'administration du traitement curatif en phase aiguë. Ces critères limitent drastiquement l'accès à ce traitement : on estime que seuls 10% des patients atteints d'un AVC en bénéficient. L'objectif de ce travail est de proposer un modèle prédictif de l'évolution de l'AVC qui permette d'identifier le volume de tissu à risque de chaque patient. Ce volume, qui correspond au bénéfice potentiel du traitement, permettra de mieux orienter le médecin dans sa décision de traiter.

Dans une première partie, nous présentons les stratégies de construction et d'évaluation des modèles prédictifs. Dans un contexte de faible prévalence, nous montrons les limites des indices issus de la courbe ROC et l'intérêt d'indices issus de la courbe précision-rappel. Ensuite, nous comparons des approches prédictives utilisant des modèles de régression ou des modèles d'apprentissage automatique ainsi que différents traitements d'images. Une méthode de prédiction volumique inspirée de l'analyse fonctionnelle est proposée et comparée au seuillage, la méthode classiquement utilisée. Enfin des critères d'éligibilité au traitement basés sur les prédictions volumiques sont proposés.

Dans la seconde partie, nous proposons un modèle de segmentation de lésion adapté aux spécificités de la séquence d'imagerie T2 FLAIR, la séquence de référence pour l'identification de la lésion finale. L'originalité du modèle de segmentation est de proposer deux niveaux de régularisation de la forme de la lésion : un niveau local pour être robuste au bruit de mesure et un niveau régional pour être robuste aux artefacts d'imagerie. Enfin, nous présentons le logiciel de gestion et de traitement d'images que nous avons développé sous la forme d'une librairie de fonctions du logiciel R. Cette librairie, nommée MRIaggr, est disponible sur le site du CRAN (<http://cran.r-project.org/>) et facilite la réutilisation des méthodes utilisées dans ce travail.

Mots clés : évaluation de biomarqueurs, modèles prédictifs, segmentation d'images, imagerie par résonance magnétique, accident vasculaire cérébral, logiciel R.

Abstract

Stroke is a serious disease that needs emergency health care. Due to potential side effects, the patients must fulfil very restrictive criteria for eligibility to the curative treatment. These criteria limit drastically the accessibility to treatment : currently, an estimated 10% of stroke patients are treated. The purpose of this work was to develop a statistical framework for stroke predictive models. These models should enable the identification of the tissue at risk in individual patients. Because the amount of tissue at risk corresponds to the potential benefit of the treatment, this identification brings crucial information for treatment allocation.

In a first part, we deal with building and assessing predictive models. We show the limits of ROC-based indexes in a low-prevalence context and we emphasize the interest of indexes based on the Precision-Recall curve. Then, we compare predictive approaches based on regression models or on machine learning methods in conjunction with various image processing methods. A functional modelling framework for volumic prediction is introduced and compared with the usual thresholding method. Finally, eligibility criteria for stroke treatment are proposed on the basis of the results given by the optimal predictive model.

In the second part, we develop a lesion segmentation model adapted to the specificities of the T2 FLAIR imaging sequence –the reference sequence for the identification of the final stroke lesion. Our contribution is to propose two levels of regularization for the lesion shape : a local level to correct for noise and a regional level to ensure robustness to image artefacts. Finally, we present the software we developed for the management and the processing of the images. This software, named MRIaggr, is a package for the R statistical environment. It is available on the CRAN website (<http://cran.r-project.org/>) and allows the use of the same methods for future works.

This document is written in French but includes four articles written in English that summarize the main results of the work :

- The ability of the area under the ROC curve (AUC) and the area under the Precision-Recall curve (AUPRC) to summarize the quality of a biomarker (e.g., the predicted lesion risk) according to the prevalence : see section 2.4 page 29.
- The comparisons between stroke predictive models and their use for defining the criteria for eligibility to stroke treatment : see section 3.5 page 60.
- The proposed lesion segmentation algorithm and its validation on simulated and real data : see section 1.6 page 125.
- The description of the main functionalities of the MRIaggr package : see section 2.4 page 178.

Articles are all independent and self-containing.

keywords : biomarker assessment, predictive modelling, image segmentation, magnetic resonance imaging, stroke, R software.

Remerciements

Mes remerciements s'adressent en premier lieu à mon directeur de thèse le Dr. Delphine Maucourt-Boulch pour m'avoir permis de découvrir et de travailler sur ce sujet passionnant qu'est la modélisation de données IRM pour l'AVC. Merci pour la confiance que tu m'as accordée et la liberté de travail dont j'ai bénéficié. Je remercie également le Pr. Norbert Nighoghossian, mon co-directeur de thèse, pour m'avoir donné l'opportunité de travailler sur la base de données I-know et pour avoir favorisé les échanges avec CREATIS. Un grand merci également au Dr. Tae Hee Cho pour m'avoir conseillé tout au long de mon travail, notamment pour la patience et la bienveillance dont tu as fait preuve lorsque je te proposais des résultats statistiques (probablement un peu obscurs) dont il fallait extraire l'intérêt clinique.

J'ai également bénéficié d'une aide précieuse (et un peu inattendue je dois dire) de la part des membres de mon comité de pilotage : M. Vivian Viallon, M. Martyn Plummer et M. Mickaël Unser. Ces deux réunions, en fin de première et seconde année, m'ont permis de dresser un bilan et de dégager des axes de travail qui, finalement, ont plutôt bien abouti. Un immense merci à tous mes collègues du service de biostatistiques des Hospices Civils de Lyon. Ce fut pour moi l'occasion d'apprendre et de partager sur des sujets très variés (statistiques mais pas que), le tout dans une ambiance très conviviale. Un merci particulier à Fabien pour toute l'aide que tu m'as apporté et pour m'avoir donné l'opportunité de faire des TP. Merci à Jean pour tes corrections mais aussi pour tes discussions, parfois mouvementées, sur l'écriture d'un article scientifique. Merci à mes co-bureaux pour tous ces moments partagés, pour avoir su trouver des solutions aux petits problèmes de la vie courante (administratifs, logiciels ou autres). Un peu moins merci pour avoir rempli le serveur et pour m'avoir embarqué dans une histoire de comparaison par paires qui m'a donné bien des sueurs froides (n'est ce pas Guénolé). Enfin merci à Emmanuelle, Laurent et Aurélien pour m'avoir remis à jour au niveau tennis et football (même si des lacunes persistent me direz vous) et à Zoé, Mad, Céren et Benjamin pour vos discussions et votre bonne humeur.

Mes remerciements vont également à mes parents et à mon frère pour leur soutien, leurs conseils (même si je ne les suis pas toujours!). Merci également pour le temps que vous avez consacré à la relecture de ma thèse et qui a permis d'éliminer bon nombre de fautes. Merci à mes colloc' et assimilées (P..... et R.....) pour tous ces bons moments. Merci à Spiro (et à ses maîtres) ainsi qu'à Aladin (que l'on a trop peu vu). Merci également à Christian, rien de mieux en effet que quelques burpees pour se remettre d'une journée de travail. Et merci à Yuka pour ta patience et ta bonne humeur : je dois avouer que certains vendredi soir les partitions se transformaient en hiéroglyphes.

Je voulais également remercier le corps enseignant qui m'a formé tout au long de ma scolarité. Une mention spéciale pour M. Leboulch et M. Secchi pour m'avoir tant appris et transmis (respectivement) le goût des maths et le goût des stats. Merci également à Mme Mironescu et M. Vial pour m'avoir aidé dans mon orientation vers les biostatistiques.

Merci enfin au ministère pour avoir financé cette thèse sans quoi rien de cela n'aurait été possible (du moins pas en trois ans).

Table des matières

Résumé	i
Abstract	ii
Remerciements	iii
Table des matières	v
Table des figures	viii
Liste des tableaux	x
Nomenclature	xi
Introduction	xiii
Partie I Contexte de l’AVC ischémique	1
1 Éléments d’anatomie du cerveau humain	1
2 Aspect médical de l’AVC ischémique	3
3 Les données d’IRM pour l’AVC	5
3.1 Principe de fonctionnement de l’IRM	6
3.2 La séquence en T1	7
3.3 La séquence en T2	8
3.4 La séquence en diffusion	9
3.5 La séquence en perfusion	10
4 Éligibilité des patients à la thrombolyse	12
Partie II Construction et évaluation de modèles prédictifs	14
1 Généralités sur la modélisation prédictive	14
1.1 Définition du modèle et de son ajustement	14
1.2 Prédiction pour une nouvelle observation	16
1.3 Évaluation d’un modèle prédictif	16
1.4 Optimisme et correction par rééchantillonnage	18
2 L’évaluation des modèles prédictifs pour l’AVC	20
2.1 Prise en compte de la structure des données dans l’évaluation	20
2.2 Indices de performance des modèles prédictifs	20

2.3	Les limites des indices issus de la courbe ROC	25
2.4	Comparaison de la pertinence de l'AUC et de l'AUPRC pour l'évaluation d'un biomarqueur	26
2.5	Critères retenus pour l'évaluation des modèles prédictifs	40
3	Comparaison de modèles prédictifs pour l'AVC	41
3.1	Les approches prédictives pour l'AVC	41
3.2	Traitement des images issues d'IRM	45
3.3	Étude préliminaire : évaluation de l'intérêt des méthodes d'apprentissage automatique	50
3.4	Améliorations proposées	51
3.5	Étude comparative des différentes améliorations proposées	53
3.6	Estimation du volume de nécrose à l'aide d'une approche fonctionnelle .	87
4	Conclusion	92
Partie III Développement d'outils d'analyse d'image		93
1	Segmentation multi-paramétrique de lésion avec régularisation de forme	93
1.1	Enjeu de la segmentation de lésion	93
1.2	Les algorithmes de segmentation d'image pour l'AVC	96
1.3	Objectif	106
1.4	Le modèle de mélange pour la segmentation d'image	107
1.5	Extension du modèle de Potts proposée	115
1.6	Estimation du modèle de segmentation	117
1.7	Vers une meilleure estimation du modèle de Potts	165
2	Développement logiciel pour l'AVC	170
2.1	Motivations	170
2.2	Présentation de la librairie de fonctions MRIaggr	171
2.3	Pipeline de pré-traitements proposé	174
2.4	Article soumis au Journal of Statistical Software	177
2.5	Nombre de téléchargements de la librairie depuis mars 2015	208
2.6	Perspectives	209
3	Conclusion	211
Conclusion générale		212
Production scientifique		214
Références		216
Annexe A : Présentation de la cohorte de patients étudiée		234
1	Critères d'inclusion des patients	234
2	Schéma de l'étude	234
3	Données disponibles	235

Annexe B : Présentation des méthodes de prédiction	237
1 La régression logistique (LR)	237
2 La régression logistique additive (ALR)	238
3 La régression logistique à effets mixtes (MLR)	240
4 La régression autologistique (autoLR)	241
5 Les arbres de classification (CART)	242
6 L'algorithme Adaboost (ADA)	243
 Annexe C : Correction des probabilités à posteriori du groupe lésion par un algorithme de croissance de régions	 245

Table des figures

I.1	Synthèse des éléments anatomiques relatifs au cerveau humain.	2
I.2	Représentation schématique de l'évolution d'un AVC	3
I.3	Représentation schématique des évolutions possibles de deux AVC suivant l'option de traitement	5
I.4	Discrétisation du volume cérébral en voxels	5
I.5	Principe de fonctionnement de l'IRM	7
I.6	Représentation des paramètres IRM anatomiques : T1 et T2	8
I.7	Représentation des paramètres IRM diagnostiques : DWI, ADC et T2 FLAIR	9
I.8	Représentation des paramètres IRM de perfusion : TTP, MTT, TMAX, CBF et CBV	11
I.9	Courbe de concentration du produit de contraste et lien avec les paramètres de perfusion	11
I.10	Impact du délai depuis les symptômes sur l'effet du traitement	13
I.11	Profils de patients proposés par l'étude DEFUSE	13
II.1	Optimisme et validation croisée	17
II.2	Représentation schématique de l'évaluation des performances du modèle prédictif	21
II.3	Comparaison de l'AUC, des AUC partielles et de l'AUPRC à différentes prévalences	39
II.4	Identification des voxels correspondant au cerveau	46
II.5	Identification de différentes structures cérébrales	46
II.6	Filtrage du paramètre IRM TMAX	48
II.7	Distribution des valeurs de paramètre IRM DWI par patient	49
II.8	Principe de l'approche fonctionnelle pour l'estimation du volume à risque	89
II.9	Comparaison entre les volumes prédits par l'approche fonctionnelle et la méthode de seuillage	91
III.1	Exemple de cartographies IRM pour trois patients à l'IRM d'admission et à l'IRM finale	95
III.2	Application d'une méthode de surface de niveaux en segmentation de lésion	102
III.3	Algorithme de croissance de régions pour la segmentation de lésion	102
III.4	Vue d'ensemble de la méthode de détection de lésion proposée par SEGHER et al., 2008	102
III.5	Application du modèle de Markov caché proposée par DWYER et al., 2008	103

III.6	Classification obtenue avec l'algorithme ISODATA (SOLTANIAN-ZADEH et al., 2007)	103
III.7	Application de l'algorithme 'Mean Shift' proposée par LI et al., 2009	103
III.8	Exemple de retraitement de la zone de lésion proposée par MEILŪNAS et al., 2003	103
III.9	Exemple de système de voisinage en deux dimensions.	109
III.10	Illustration du chien modèle	114
III.11	Exemple d'application de l'algorithme "Mean Shift"	114
III.12	Méthodes de calcul du potentiel local et régional	116
III.13	Schéma récapitulatif des différentes étapes de l'algorithme EM pour l'estimation du modèle de mélange	122
III.14	Groupes formés par le modèle de mélange	123
III.15	Cartographies T2 FLAIR de 5 patients ayant une maladie de la substance blanche marquée	163
III.16	Exemple de résolution globale du modèle de Potts	169
III.17	Exemple de masques cérébraux obtenus avec la librairie MRIaggr pour 3 patients	176
III.18	Nombre de voxels de la lésion finale hors du masque cérébral	177
III.19	Exemple de contraste insuffisant pour l'identification du type de tissu	177
III.20	Nombre de téléchargements hebdomadaires de la librairie MRIaggr	208
III.21	Exemple de sortie logiciel pouvant aider le médecin dans sa prise de décision .	210
B.1	Exemple d'arbre de décision	243
B.2	Schéma de fonctionnement de l'algorithme Adaboost	244
C.1	Étapes de correction des probabilités a posteriori d'appartenance au groupe lésion sur un exemple fictif	246

Liste des tableaux

II.1	Modèles prédictifs utilisés dans le manuscrit	17
II.2	Indices d'évaluation des approches prédictives utilisés dans la littérature . . .	21
II.3	Table de contingence	23
II.4	Synthèse des approches prédictives voxel à voxel proposées dans la littérature	44
II.5	Étapes de normalisation associées à chaque paramètre	48
II.6	Résultats de l'ANOVA sur les performances du modèle prédictif avec un effet patient, un effet modèle et un effet filtrage	58
III.1	Méthodes de segmentation pour l'AVC	99
III.2	Synthèse des libraires R existantes pour l'analyse de données d'imagerie pour l'AVC	172
A.1	Instants de prise en charge et de mesure IRM selon les patients.	235
A.2	Variables descriptives de la cohorte étudiée	236

Nomenclature

Abréviations :

Les abréviations utilisées dans le manuscrit seront les abréviations anglaises par cohérence avec les notations utilisées dans les articles, excepté pour les abréviations IRM, %*BC* et v.a. .

% <i>BC</i>		Pourcentage d'observations bien classées
<i>ADA</i>	Adaptive boosting	Adaboost ou algorithme de stimulation adaptative
<i>ADC</i>	Apparent Diffusion Coefficient	Coefficient de diffusion apparent
<i>ALR</i>	Additive Logistic Regression	Régression logistique additive
<i>ANOVA</i>	Analysis Of Variance	Analyse de la variance
<i>autoLR</i>	Autologistic Regression	Régression autologistique
<i>AUC</i>	Area Under the ROC curve	Aire sous la courbe ROC
<i>AUPRC</i>	Area Under the PR curve	Aire sous la courbe PR
<i>CBF</i>	Cerebral Blood Flow	Débit sanguin cérébral
<i>CBV</i>	Cerebral Blood Volume	Volume sanguin cérébral
<i>CSF</i>	Cerebral Spinal Fluid	Liquide cérébro-spinal
<i>DWI</i>	Diffusion-Weighted Imaging	Imagerie pondérée en diffusion
<i>EM</i>	Expectation-Maximization algorithm	Algorithme espérance-maximisation
<i>FLAIR</i>	Fluid-Attenuated Inversion Recovery	Séquence en inversion-récupération (élimine le signal provenant des fluides)
<i>GM</i>	Grey Matter	Substance grise
<i>GRE</i>	Gradient-Recalled Echo	Séquence pondérée en écho de gradient
<i>IRM</i>	Magnetic Resonance Imaging (<i>MRI</i>)	Imagerie par résonance magnétique
<i>LR</i>	Logistic Regression	Régression logistique
<i>MLR</i>	Mixed Logistic Regression	Régression logistique à effets mixtes
<i>MTT</i>	Mean Transit Time	Temps de transit moyen
<i>NIHSS</i>	National Institute of Health Stroke Scale	Score clinique de l'AVC des instituts américains de la santé
<i>NPV</i>	Negative Predictive Value	Valeur prédictive négative
<i>PPV</i>	Positive Predictive Value	Valeur prédictive positive
<i>PR</i>	Precision Recall	Précision et rappel
<i>ROC</i>	Receiver Operating Characteristic	Fonction d'efficacité du récepteur

$rt - PA$	recombinant tissue Plasminogen Activator	Altéplase, forme recombinante de l'activateur tissulaire du plasminogène
Se	Sensitivity	Sensibilité
Sp	Specificity	Spécificité
$TMAX$	Time-to-maximum	Temps de maximum
TTP	Time To Peak	Temps de pic
v.a.		Variable aléatoire
WM	White Matter	Substance blanche

Notations :

t	transposée d'un vecteur ou d'une matrice
\mathcal{C}	matrice de design contenant les variables cliniques (en colonne) pour chaque patient (en lignes)
\mathcal{X}	matrice de design contenant les variables d'imagerie (en colonne) pour chaque voxel (en lignes)
i	indice de voxel
id	indice de patient
n	nombre de voxels
$\mathbf{N}(i)$	ensemble des voxels voisins du voxel i
ξ	indicateur d'appartenance
$f_{\mathcal{G}}$	densité de la loi normale.
$\mathbb{E}[X]$	espérance de la v.a. X .
$\mathbb{E}[X Y]$	espérance de la v.a. X conditionnellement à la v.a. Y .
$\mathbb{P}[A]$	probabilité d'observer l'évènement A .
$\mathbb{V}ar[X]$	variance de la v.a. X .
$ \hat{V} - V $	erreur volumique en valeur absolue (exprimée en mL)
$\rho(V, \hat{V})$	corrélacion entre le volume observé et le volume prédit

Les variables aléatoires seront désignées par des lettres majuscules (ex : X) et leur réalisation par des lettres minuscules (ex : x).

La matrice de design sans indice (ex : \mathcal{X}) correspond à la matrice de design pour l'ensemble des voxels et des patients. L'indication par id ou i restreint au patient id ou au voxel i (ex : \mathcal{X}_{id}). Le double indication par id et i restreint au voxel i du patient id (ex : $\mathcal{X}_{i,id}$).

Introduction

L'Accident Vasculaire Cérébral (AVC) est un problème de santé publique majeur pour les pays occidentaux. Il s'agit d'une maladie lourde de conséquences puisque près de 15% des patients décèdent au terme du premier mois et 75% des survivants garderont un handicap permanent. Avec 130 000 nouveaux cas par an en France, soit une incidence d'environ 2 pour mille personnes années, l'AVC est la première cause de handicap acquis chez l'adulte, la seconde cause de démence et la troisième cause de mortalité (HAUTE AUTORITÉ DE SANTÉ, 2009). Même si l'AVC touche principalement les personnes âgées (âge moyen de survenu de 73 ans), une augmentation préoccupante de l'incidence a été reportée chez les moins de 55 ans (BÉJOT et al., 2014 ; KISSELA et al., 2012). Parmi les principaux facteurs de risque de l'AVC on retrouve donc l'âge mais aussi l'hypertension, le diabète, le tabagisme ou encore l'arythmie - voir l'article d'ALLEN et al., 2008, pour une revue plus complète. L'AVC est causé par l'obstruction d'un vaisseau sanguin (AVC ischémique, 80% des cas) ou par la rupture du vaisseau sanguin (AVC hémorragique, 20% des cas). Dans les deux cas, il conduit à une diminution brutale de l'irrigation sanguine pour les tissus en aval du vaisseau. Les réserves métaboliques des neurones étant limitées (environ 3 minutes en autonomie complète), seule une réirrigation précoce permettra de préserver le tissu cérébral.

La démonstration en 1995 de l'efficacité de la thrombolyse pour traiter l'AVC ischémique (THE NINDS GROUP, 1995) a profondément modifié la prise en charge de l'AVC. Ce traitement, s'il est administré suffisamment tôt, permet une réduction du handicap dans 30 % des cas pour une augmentation de 6% des complications hémorragiques. Une des principales limites à l'utilisation de la thrombolyse est la diminution de son efficacité lorsqu'elle est administrée tardivement. En moyenne, les patients arrivant plus de 4h30 après le début des symptômes ou ignorant l'heure de début des symptômes ne vont pas retirer de bénéfice à être traités par thrombolyse (ATLANTIS, ECASS, AND NINDS GROUPS, 2004). Ainsi en France, le pourcentage de patients ayant été thrombolysés suite à un AVC est estimé à 1% en 2006 (1080 pour 80 000 hospitalisations, BARDET, 2007) et aux alentours de 10% aujourd'hui. Malgré des progrès importants réalisés pour accélérer la prise en charge, la fenêtre de 4h30 reste toujours très contraignante et peu adaptée puisque la cinétique d'expansion de la lésion engendrée par l'AVC est propre à chaque patient (SAVER, 2006).

De grands espoirs se fondent sur l'utilisation de l'imagerie médicale afin de pouvoir quantifier le bénéfice potentiel du traitement pour chaque patient. Au cours des vingt

dernières années, les progrès de l'imagerie médicale ont en effet révolutionné la prise en charge de l'AVC. En permettant une visualisation en 3 dimensions de l'ensemble du cerveau elle rend possible une détection fiable et peu invasive des lésions cérébrales. En plus de son intérêt diagnostique, l'imagerie médicale pourrait avoir un intérêt pronostique via l'identification des zones d'évolution future de l'AVC. Cette information permettrait une meilleure allocation du traitement et donc une meilleure prise en charge de l'AVC. Trois techniques d'imagerie médicale peuvent être utilisées pour étudier l'AVC : la tomодensitométrie, l'imagerie par résonance magnétique (IRM) et la tomographie par émission de positons. Dans le cadre de cette thèse, nous nous focaliserons sur l'IRM, qui est l'outil de référence en pratique clinique (JAUCH et al., 2013; HAUTE AUTORITÉ DE SANTÉ, 2009) même si la tomодensitométrie, moins coûteuse, est extrêmement répandue. Les méthodes proposées dans cette thèse n'étant pas spécifiques à l'IRM, elles devraient aussi être applicables en tomодensitométrie.

L'IRM permet de mesurer des grandeurs physiques caractéristiques des zones observées en utilisant des champs magnétiques. Le principe est de perturber l'orientation magnétique des protons H^+ à l'aide d'une onde électromagnétique afin de pouvoir mesurer des grandeurs caractéristiques de leur retour à l'équilibre. Les spécificités du milieu vont influencer le temps de retour à l'équilibre des protons : il sera par exemple plus court pour les solides que pour les liquides. Ainsi les variations de mobilité des protons suite à l'œdème généré par l'AVC vont permettre de détecter les lésions. Suivant les paramètres d'acquisition de l'IRM il est possible de mesurer différentes grandeurs ou bien de détecter l'arrivée d'un produit de contraste. Ces paramètres d'acquisition ont été standardisés en séquences IRM, chacune permettant de calculer un ou plusieurs paramètres caractéristiques du phénomène étudié. Ainsi la séquence en diffusion, très sensible aux différences de mobilité de l'eau, permet de détecter dès les premiers signes la présence d'un AVC. Passé les premiers jours, cette séquence perd en sensibilité et une autre séquence, la séquence T2 FLAIR, lui est préférée pour évaluer le volume de lésion final causé par l'AVC. La séquence en perfusion permet d'évaluer la qualité de l'irrigation sanguine en mesurant le retard d'arrivée au cerveau d'un produit de contraste injecté au patient. Enfin des séquences IRM anatomiques comme le T1 permettent de repérer les structures cérébrales (boite crânienne, substance blanche, substance grise, ...).

Du point de vue diagnostique, l'IRM a rempli les attentes puisque la séquence en diffusion permet d'établir de façon fiable la présence d'un AVC et une autre séquence, la séquence en T2*, permet d'identifier d'éventuelles hémorragies (CHALELA et al., 2007; SCHELLINGER et al., 2010). En revanche l'établissement d'un pronostic fiable de l'AVC est nettement plus complexe. D'une part, certaines lésions observées en diffusion semblent réversibles (FIEHLER et al., 2002; OLIVOT et al., 2009b) même si l'importance du phénomène est toujours contesté (RINGER et al., 2001; CAMPBELL et al., 2012). D'autre part, même si l'intérêt pronostique de la séquence en perfusion a été établi il y a plus de 10 ans (THIJS et al., 2001; PARSONS et al., 2001; SHIH et al., 2003), aucun des paramètres proposés par cette séquence n'est apparu être un indicateur fiable de l'évolution future de l'AVC (KIDWELL et al., 2003). Malgré de nombreuses études sur le sujet (pour un aperçu voir le tableau 1 de l'article de RIVERS et al., 2006, aucun consensus n'a

été trouvé sur l'utilisation cette séquence. La modélisation statistique apparaît aujourd'hui être une approche prometteuse qui permettrait de combiner l'information fournie par les différents paramètres IRM.

Cette approche initiée par l'équipe d'Ona Wu en 2001 (WU et al., 2001) trouve aujourd'hui un large écho avec de nombreux modèles prédictifs publiés. Différentes équipes ont ainsi proposé des modèles variés : régression logistique (WU et al., 2006), régression logistique à effets mixtes (KIDWELL et al., 2013), régression logistique additive (BOUTS et al., 2013), réseaux de neurones artificiels (HUANG et al., 2010 ; BAGHER-EBADIAN et al., 2011), machines à vecteurs de support (HUANG et al., 2011), régression spectrale (SCALZO et al., 2012), forêts d'arbres décisionnels (BOUTS et al., 2013). Même si ces modèles ont montré des performances intéressantes, il manque dans la littérature des études comparant l'ensemble des méthodes prédictives.


De plus, l'application des méthodes statistiques usuelles sur des données d'imagerie médicale se heurte à un certain nombre de difficultés. Comme dans de nombreux domaines, les données sont altérées par du bruit de mesure, ici imputable aux imperfections du champ magnétique ou aux instabilités des méthodes de traitement du signal. Plus problématique, le signal IRM n'est pas toujours spécifique de l'AVC. La séquence T2 FLAIR est par exemple à la fois sensible à la maladie de la substance blanche et à l'AVC. Concernant la séquence en perfusion, des valeurs élevées de retard sont souvent observées à proximité du liquide céphalo-rachidien alors qu'elles devraient uniquement marquer les zones en déficit d'irrigation. Par ailleurs, malgré les étapes de recalage des données, des décalages spatiaux subsistent parfois entre les mesures effectuées à différents temps chez un même patient. Enfin, l'AVC évolue avec ses propres spécificités chez chaque patient : type de vaisseau atteint, présence d'un réseau d'irrigation secondaire, antécédents médicaux du patient, etc. Tous ces éléments peuvent être des sources d'erreurs pour le modèle prédictif. Une validation minutieuse du modèle et de son intérêt clinique doit donc être effectuée.

La fiabilité d'un modèle prédictif est aussi liée à la fiabilité du jeu de données d'apprentissage, c'est-à-dire à la définition de la lésion finale chez chaque patient. L'œil humain étant particulièrement performant pour distinguer des formes sur des images, cette étape est généralement effectuée par un médecin. Il s'agit d'un travail fastidieux, et ce d'autant plus que l'on souhaite un jeu de données conséquent, mais c'est aussi un travail délicat. Comme évoqué précédemment, la lésion causée par l'AVC ne se détache pas toujours nettement du reste du tissu cérébral. Aussi il peut être intéressant de regarder non pas un mais plusieurs paramètres IRM, et de prendre en compte la structure spatiale de la lésion pour mieux la délimiter. De nombreuses méthodes de segmentation ont été proposées pour remplacer ou assister le travail du médecin : en se restreignant aux méthodes évaluées dans la littérature sur des patients atteints d'AVC, on en compte actuellement plus de 40. Ces méthodes reposent sur différentes approches : modèles de déformation de surface (par exemple : STEIN et al., 2001 ; WEINMAN et al., 2003 ; MOURIDSEN et al., 2013), modèles d'apprentissage (par exemple : USINSKAS et al., 2003), modèles de partitionnement de

données (par exemple : JACOBS et al., 2000 ; FORBES et al., 2010b ; ARTZI et al., 2013) et méthode de détection de données anormales (par exemple : MALDJIAN et al., 2001 ; SEGHER et al., 2008 ; MAH et al., 2014). Une limite commune à ces approches est qu'elles sont sensibles aux artéfacts d'imagerie. De tels artéfacts sont fréquents sur les images T2 FLAIR notamment à cause de la maladie de la substance blanche. Cette maladie généralement présente chez les patients âgés est caractérisée par un signal IRM en T2 FLAIR similaire à celui de l'AVC. L'algorithme de segmentation risque donc de confondre ces deux maladies entraînant une mauvaise identification de la zone touchée par l'AVC.

L'objectif de la thèse est de développer les outils nécessaires à la mise en œuvre de modèles prédictifs, puis de les appliquer sur des données réelles afin de proposer une caractérisation précise de la zone et du volume de tissu menacé par l'AVC. Les données utilisées proviennent de l'étude I-know (ØSTERGAARD, 2006) : une étude européenne multicentrique ayant pour objectif une meilleure compréhension de l'AVC en vue d'améliorer sa prise en charge. Seule une partie des patients inclus dans l'étude ont été thrombolysés mais tous ont passé des IRM à différents temps afin de pouvoir mesurer les évolutions précoces et tardives de l'AVC. Une description plus détaillée des patients étudiés est disponible dans l'annexe A.

Le premier chapitre de ce mémoire introduit les notions essentielles de l'AVC et de l'IRM pour l'AVC. Un lecteur familier avec ce contexte d'étude pourra directement passer à la suite du mémoire, qui est divisé en deux parties :

- La première partie est centrée sur **la construction et l'évaluation des modèles prédictifs**. Nous montrerons tout d'abord que les outils classiques d'évaluation de modèles ne sont pas adaptés pour les données d'imagerie et nous proposerons des outils alternatifs. Puis nous comparerons les principaux modèles prédictifs proposés dans la littérature et nous proposerons diverses améliorations afin de retenir un modèle optimal. Nous nous intéresserons à la pertinence clinique de ce modèle, notamment sa capacité à prédire la croissance de l'AVC ainsi que son utilisation pour l'identification des patients susceptibles de répondre au traitement. Enfin, nous présenterons une méthode originale pour effectuer une prédiction volumique à partir des risques de nécrose prédits permettant de tenir compte des caractéristiques cliniques du patient.
- La seconde partie présente **des outils de traitement d'images pour l'AVC**. Nous commencerons par proposer un algorithme de segmentation automatique de lésion pour l'AVC, robuste au bruit et aux artéfacts présents sur les cartographies T2 FLAIR. Cet algorithme sera évalué sur données simulées et données réelles. L'approximation du champ moyen étant utilisée pour estimer certains des paramètres du modèle, l'utilisation d'une approche alternative récemment publiée (STORATH et al., 2014) sera envisagée. Enfin nous présenterons le logiciel de gestion et de traitement d'images qui a été développé au cours de cette thèse. Il regroupe, sous la forme d'une librairie du logiciel , les fonctions qui ont été utilisées pour traiter les données d'imagerie, évaluer les modèles et visualiser les résultats.

Partie I

Contexte de l'AVC ischémique

1 Éléments d'anatomie du cerveau humain

Le cerveau est le centre du système nerveux humain : d'un volume de l'ordre 1 200 mL, il concentre la plus grande partie des neurones (environ 90 milliards). Protégé par la boîte crânienne, il est composé de deux hémisphères quasi-symétriques subdivisables en zones fonctionnelles sensorielles, motrices et associatives. Les deux hémisphères sont séparés par un profond sillon appelé scissure interhémisphérique. Celui-ci étant quasi-linéaire, il est généralement modélisé par un plan, le **plan sagittal médian**. La quasi-symétrie des hémisphères permet d'assimiler ce plan à un plan de symétrie pour le cerveau. En imagerie cérébrale, les hémisphères sont représentés de façon inversée (convention radiologique) : l'hémisphère droit est représenté à gauche sur les images issues de l'IRM.

Les deux hémisphères cérébraux sont reliés par des fibres nerveuses mais conservent une certaine autonomie, notamment en terme d'irrigation sanguine. Le réseau d'irrigation cérébral est approvisionné par quatre artères : les artères carotides internes (gauche et droite) et les artères vertébrales (gauche et droite). Ces artères se rejoignent au polygone de Willis pour former six artères : les artères cérébrales antérieures (gauche et droite), les artères cérébrales moyennes (gauche et droite) et les artères cérébrales postérieures (gauche et droite). Chaque hémisphère est donc desservi par trois artères qui lui sont propres. Elles vont se subdiviser en vaisseaux sanguins de plus petite taille afin d'alimenter le tissu cérébral (figure [I.1](#)).

Le tissu cérébral est composé de trois principaux types de cellules : les neurones qui traitent et transmettent des influx nerveux, les cellules gliales qui servent essentiellement de support (nutrition, structuration, protection) et les cellules souches qui assurent le renouvellement (partiel) des cellules cérébrales. Les neurones sont constitués d'un corps contenant ses principaux organites, de dendrites conduisant le signal électrique provenant d'autres neurones, d'un axone conduisant le signal vers d'autres neurones et de synapses permettant la transmission du signal entre deux neurones.

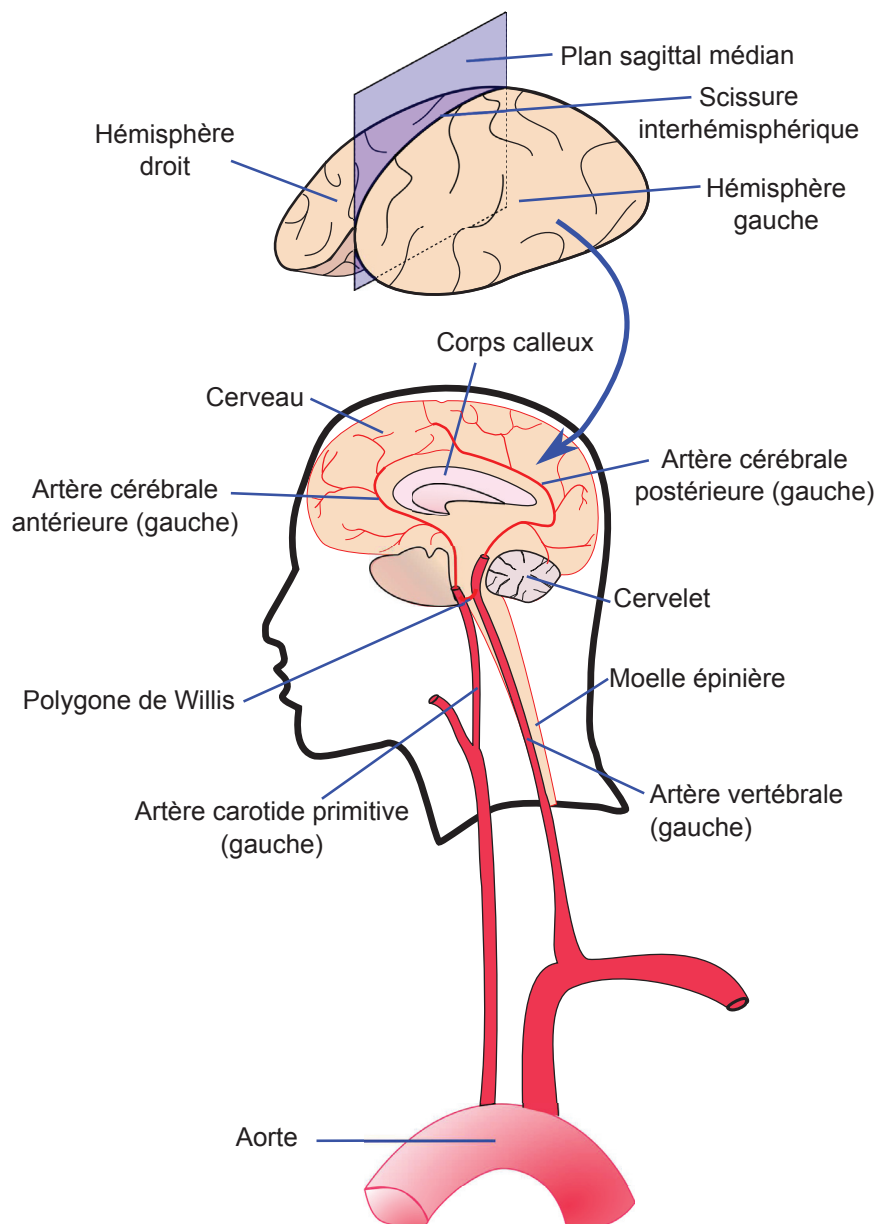


FIGURE I.1 – Synthèse des éléments anatomiques relatifs au cerveau humain.

Le tissu cérébral est souvent décomposé en deux types de tissus : la **substance grise** contenant essentiellement le corps des neurones et la **substance blanche** majoritairement composée d'axones. Ces deux tissus sont facilement distinguables, même à l'oeil nu. Par ailleurs, l'ensemble du tissu cérébral baigne dans un liquide appelé **liquide cérobro-spinal**. Ce liquide a principalement un rôle protecteur en amortissant par exemple les chocs entre le cerveau et la boîte crânienne.

La substance blanche située dans le cerveau profond est principalement irriguée par les terminaisons des artères. La petite taille des artères ainsi que l'absence de redondance des sources d'irrigation rendent la substance blanche profonde particulièrement vulnérable (HÂNCU et al., 2012) : elle est souvent endommagée chez les patients âgés. Cette détérioration de la substance blanche, aussi appelée leucoaraïose ou **maladie de la substance blanche**, est généralement symétrique et située le long du liquide cérobro-spinal. Les causes de cette

maladie ne sont pas complètement comprises. Sont évoquées dans la littérature la dégradation de l'irrigation sanguine éventuellement aggravée par la présence de plaques de lipides sur les parois des vaisseaux sanguins (athérosclérose, HÂNCU et al., 2012). En effet, cette dernière limite la vasoconstriction des vaisseaux et donc la régulation sanguine. Sont aussi évoquées une perméabilité anormale du tissu cérébral au liquide cérébro-spinal ainsi que des problèmes de circulation veineuse. Les patients âgés sont particulièrement affectés par cette maladie : chez des sujets sains, le volume de leucoaraiose est de l'ordre de 2 mL avant 60 ans, 3 mL entre 60 et 75 ans et près de 8 mL après 75 ans (VERNOOIJ et al., 2007). La leucoaraiose étant facteur de risque pour l'AVC (VERMEER et al., 2003), les patients atteints d'un AVC montrent fréquemment des signes de leucoaraiose avec des volumes pouvant aller jusqu'à une centaine de millilitres.

2 Aspect médical de l'AVC ischémique

Un AVC ischémique résulte de l'obstruction d'un vaisseau sanguin par du matériel circulant dans le sang (embole ou thrombus). Cette obstruction peut être provoquée ou aggravée par la présence d'athérosclérose, par la dissection du vaisseau ou par la compression du vaisseau exercée par une tumeur. L'obstruction entraîne une diminution brutale d'irrigation pour le tissu cérébral en aval. L'irrigation normale du tissu cérébral correspond à un débit sanguin cérébral de l'ordre de 58 mL par minute et pour 100g de tissu. En dessous de 30 mL/min/100g de tissu, le neurone n'est plus capable d'assurer sa fonction ce qui entraîne un handicap fonctionnel pour le patient. En dessous de 12 mL/min/100g de tissu, les fonctions vitales du neurone sont menacées : sans restauration rapide de l'irrigation sanguine le neurone va mourir. Ainsi, l'impact de l'AVC sur le tissu cérébral peut être schématisé en 3 zones (figure I.2) :

- **Une zone de nécrose** où les neurones sont déjà morts.
- **Une zone de pénombre** où le devenir des neurones est incertain. Encore viables, ils mourront si l'irrigation sanguine n'est pas restaurée.
- **Une zone d'oligémie** où les neurones sont affectés de manière bénigne par l'AVC. Ils retourneront à leur état normal en fin d'AVC.

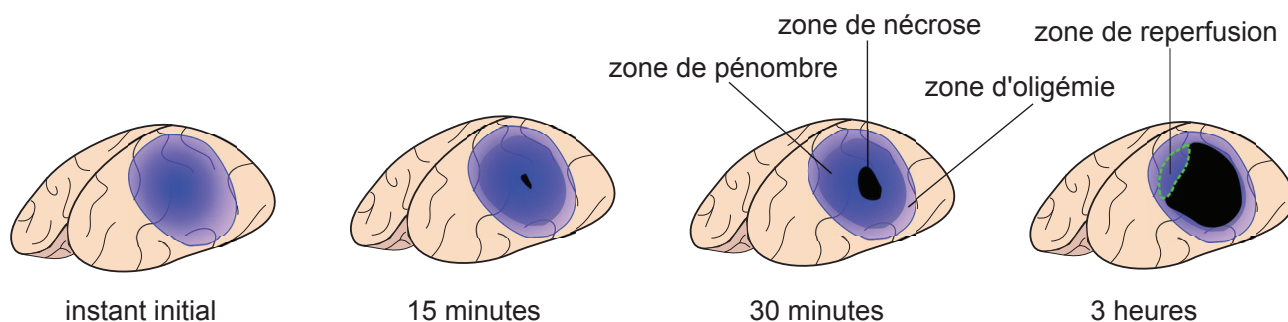


FIGURE I.2 – Représentation schématisée de l'évolution d'un AVC. La zone de nécrose se propage sur la zone de pénombre, excepté pour la zone qui a été réirriguée (zone de reperfusion représentée en pointillés verts sur le schéma).

La dégradation du tissu neuronal suite à l'AVC est un phénomène complexe (KANEKAR et al., 2012) parfois appelé cascade ischémique (BEAUCHAMP et al., 1998). Plusieurs mécanismes vont dégrader l'état de la cellule suite au manque d'irrigation sanguine (carence en adénosine triphosphate, dérèglement des pompes ioniques, émission de radicaux libres, ...). Il en résulte un déplacement d'eau vers l'intérieur des cellules, provoquant leur gonflement et la restriction de l'espace intercellulaire. C'est l'œdème cytotoxique. Cet état peut durer quelques jours jusqu'à la rupture de la structure des cellules. Le liquide intracellulaire est alors relâché dans l'espace extracellulaire pour former l'œdème vasogénique. Après plusieurs semaines, l'œdème vasogénique se résorbe et les tissus nécrosés sont éliminés.

Les conséquences d'un AVC sont très variables : du handicap mineur jusqu'au décès du patient. Afin d'évaluer l'état du patient tout au long de sa prise en charge, différents scores cliniques ont été établis. Il s'agit d'échelles de handicap croissant (0 : sans handicap, score maximum : patient décédé). Nous nous concentrerons sur un seul score clinique : le NIHSS (National Institutes of Health Stroke Scale - BROTT et al., 1989) qui évalue le déficit neurologique du patient en rubriques (vision, motricité, langage ...). Chaque rubrique est évaluée entre 0 et 3. La somme de ces items permet d'estimer l'importance du handicap du patient entre 0 et 36 : un score inférieur à 4 indique un handicap mineur tandis qu'un score supérieur à 20 indique un handicap lourd.

Lors de l'admission du patient à l'hôpital, l'objectif clinique est de stopper l'expansion de l'AVC en rétablissant une irrigation normale dans la zone de pénombre. Il s'agit donc d'effectuer une thrombolyse c'est-à-dire de dissoudre le matériel qui obstrue le ou les vaisseau(x) sanguin(s). La zone ainsi réirriguée est appelée **zone de reperfusion**. Cette opération n'est pas sans danger car les cellules endothéliales qui composent les vaisseaux sont fragilisées par le déficit d'irrigation (ZOPPO et al., 2000). Le rétablissement brutal de l'irrigation peut entraîner une rupture du vaisseau sanguin qui conduira à une hémorragie cérébrale. Ainsi, le ratio bénéfice-risque sera en faveur du traitement seulement s'il reste des tissus à risque de nécrose pouvant encore être sauvés. Or l'AVC est une maladie à cinétique rapide : la durée moyenne d'un AVC est de l'ordre de 10 heures avec une variabilité de plusieurs heures (SAVER, 2006). La zone de pénombre, qui agrège l'ensemble des tissus à risque, va rapidement évoluer vers une zone de nécrose. L'enjeu actuel est d'identifier à l'IRM d'admission la zone de pénombre restante pour chaque patient et d'établir deux profils de patients : un profil "mismatch" avec une zone de pénombre étendue et un profil "no-mismatch" où la zone de pénombre est négligeable (voir figure I.3). Le profil du patient pourra alors aider le médecin dans sa décision d'allocation du traitement.

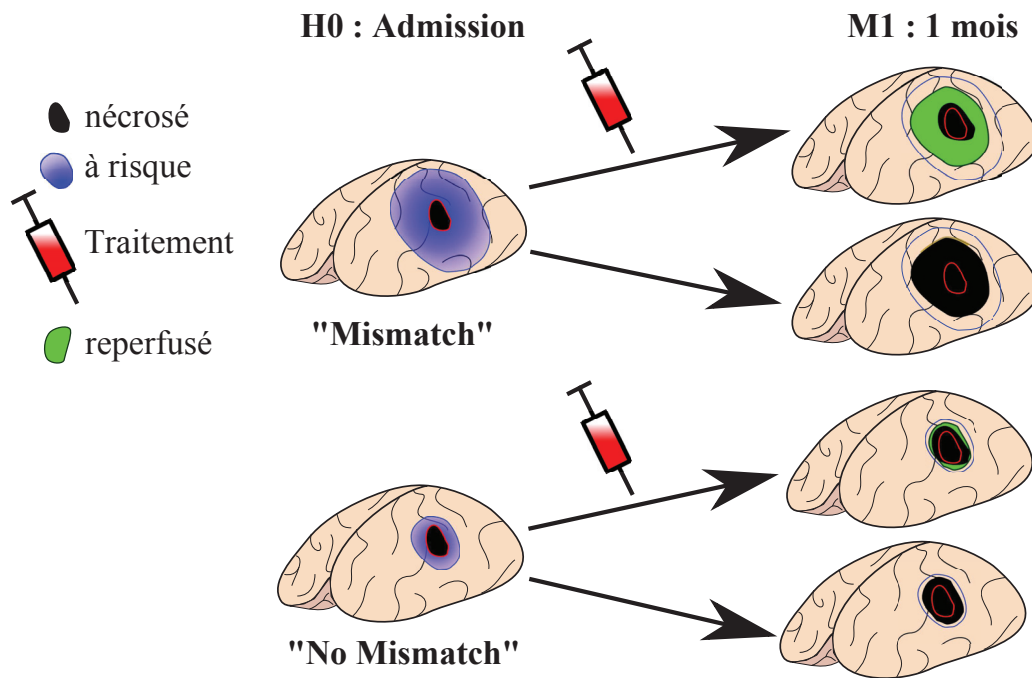


FIGURE I.3 – Représentation schématique des évolutions possibles de deux AVC suivant l'option de traitement. Le mot "Mismatch" (en français décalage) se réfère au décalage qu'il y a entre la zone de pénombre initiale et la zone de nécrose à l'instant d'observation.

3 Les données d'IRM pour l'AVC

L'identification des zones de nécrose et de pénombre repose sur les informations récoltées au cours de l'IRM du patient. Un examen par IRM permet en effet de caractériser par plusieurs grandeurs, appelées paramètres IRM, le cerveau. Pour cela, le volume cérébral est partitionné en un nombre fini de parallélépipèdes identiques appelés voxels (voir figure I.4).

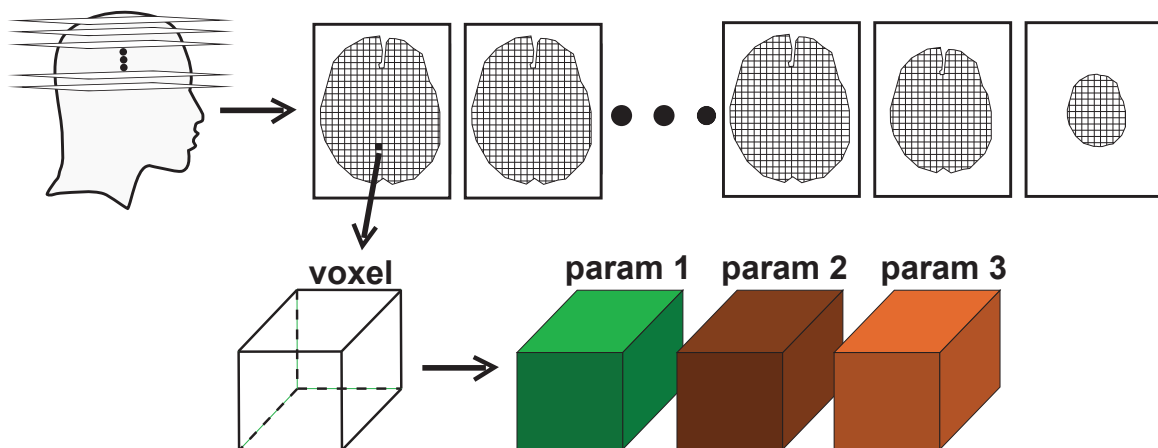


FIGURE I.4 – Discretisation du volume cérébral en différentes coupes, elles-mêmes composées de pixels. L'association des coupes permet de former des volumes élémentaires, appelés voxels, qui vont représenter le cerveau. Suite aux acquisitions d'IRM, chaque voxel est caractérisé par différents paramètres de l'IRM.

Dans notre cas les voxels mesureront 6 mm de haut et 1,875 mm de long et de large ce qui correspond à une résolution de 128×128 voxels dans le plan et de l'ordre de 20 coupes en hauteur. Plusieurs séquences IRM permettent de mesurer les valeurs de différents paramètres caractéristiques des tissus étudiés, chaque valeur pouvant être assimilée à un niveau d'intensité. Les paragraphes suivants rappellent le principe général de l'IRM et présentent les séquences d'IRM que nous utiliserons. De plus amples détails concernant les phénomènes physiques mis en jeu ainsi que les réglages spécifiques à l'acquisition de chaque séquence se trouvent dans l'ouvrage d'HOA et al., 2008.

3.1 Principe de fonctionnement de l'IRM

L'IRM permet de mesurer la réponse d'un atome à une onde électromagnétique en différents points de l'espace. L'atome d'hydrogène est généralement utilisé car il possède un moment magnétique de spin et il est très abondant dans le corps humain. Le moment magnétique de spin résulte de la rotation des charges électriques de l'atome et des propriétés quantiques intrinsèques à l'atome qui génèrent un champ magnétique, bien que de faible intensité. Il est représenté schématiquement par une flèche dont l'orientation donne le sens du champ, et la longueur donne l'intensité du champ (figure I.5).

En l'absence d'excitation externe, l'orientation des moments est aléatoire et la résultante est donc nulle. Le moment magnétique d'un atome va interagir avec le milieu : lorsqu'un atome se trouve dans un champ magnétique, il a tendance à se positionner de façon que son moment soit dans la direction du champ magnétique. Il est possible de modifier l'orientation du moment magnétique à l'aide d'une onde électromagnétique ayant une fréquence bien précise (la fréquence de Larmor). Cette dernière dépend notamment de l'intensité du champ magnétique dans lequel est plongée la particule.

Afin de pouvoir étudier la réponse des moments magnétiques à une perturbation de son orientation, l'IRM va dans un premier temps aligner les moments magnétiques à l'aide d'un champ magnétique intense. Tous les moments n'ont donc plus qu'une composante longitudinale ; leur composante transversale étant nulle. Puis des gradients de champ sont appliqués pour que le champ magnétique ait une intensité spécifique dans chaque volume élémentaire. Chaque volume contient donc des particules dont la fréquence de Larmor lui est propre. Grâce à une onde électromagnétique de fréquence spécifique (la fréquence de Larmor du volume), il est alors possible de faire basculer l'orientation des moments magnétiques de spin : transfert de la composante longitudinale en composante transversale. Après cette bascule, la vitesse de décroissance de la composante longitudinale et la vitesse de croissance de la composante transversale varient selon la composition du volume élémentaire. Ces vitesses vont donc permettre de caractériser le volume. En effet, chaque composante induit un courant électrique (loi de Faraday) dont on peut mesurer le potentiel. Cette mesure se fait typiquement à un temps fixé après l'excitation (appelé temps d'écho).

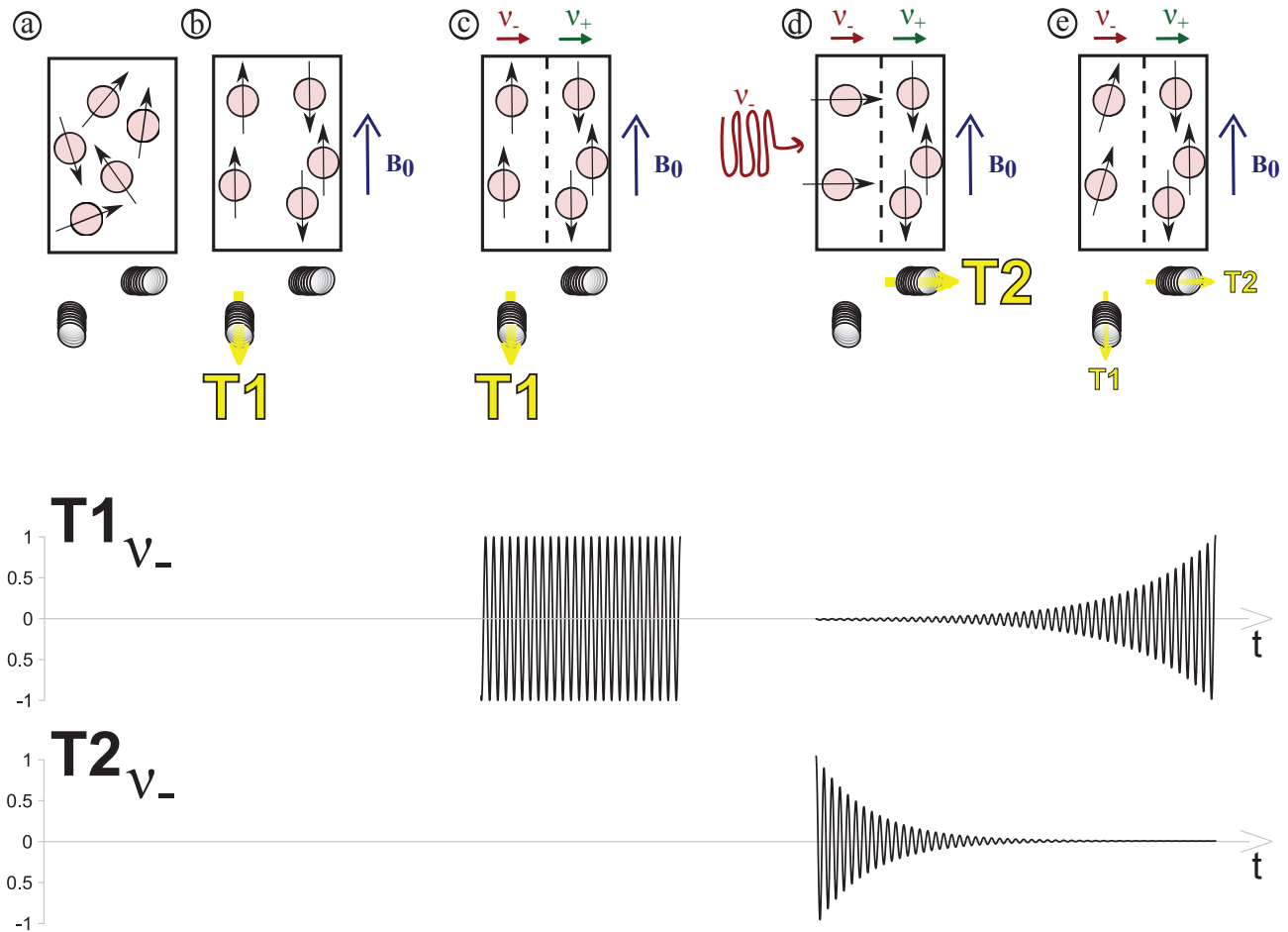


FIGURE I.5 – Principe de fonctionnement de l'IRM : (a) moment magnétique de spin au repos. (b) alignement des moments avec le champ magnétique. (c) utilisation de gradients pour modifier la fréquence de Larmor. (d) excitation à l'aide d'une onde électromagnétique. (e) décroissance de la composante transversale et croissance de la composante longitudinale. Le courant induit par la composante transversale (T2) et par la composante longitudinale (T1) du moment magnétique est représenté par une flèche jaune passant dans une bobine. L'évolution du potentiel oscillant à la fréquence de Larmor du volume considéré est représentée sur les deux graphiques du bas.

Un signal faible (hyposignal) est caractéristique d'une croissance lente ou d'une décroissance rapide tandis qu'un signal fort (**hypersignal**) est synonyme d'une croissance rapide ou d'une décroissance lente.

3.2 La séquence en T1

La croissance de la composante longitudinale est assimilable à une courbe exponentielle. Sa cinétique dépend de la mobilité des molécules à laquelle sont associés les atomes d'hydrogène (relaxation spin-réseau). De cette courbe est extrait le temps T1 correspondant au temps nécessaire pour atteindre 63% de l'aimantation longitudinale initiale. Ce temps est important lorsque la mobilité des particules est élevée (exemple : 3000 ms dans l'eau) et faible dans les tissus denses ou les grosses molécules (exemple : 810 ms pour la substance grise).

Cette séquence est principalement une séquence anatomique puisqu'elle caractérise la structure sous-jacente à chaque élément de volume. Elle fournit un contraste suffisant pour distinguer trois types anatomiques (figure I.6) : le liquide cérébro-spinal, la substance blanche et la substance grise.

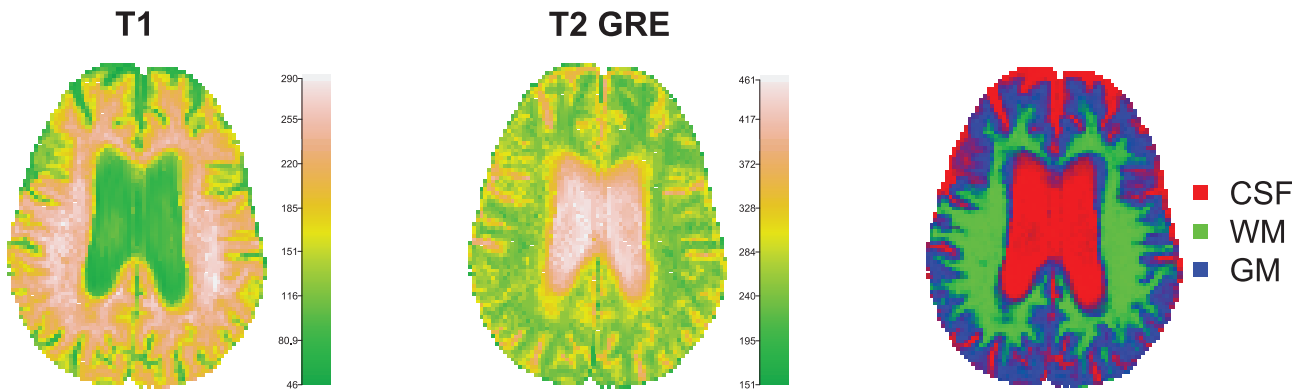


FIGURE I.6 – Paramètres anatomiques : séquence IRM pondérée T1 (à gauche) et pondérée T2 en écho de gradient (au milieu). À droite, structures anatomiques identifiées sur le T1 : liquide cérébro-spinal (CSF, en rouge), substance blanche (WM, en vert) et substance grise (GM, en bleu).

3.3 La séquence en T2

La décroissance de la composante transversale est aussi assimilable à une courbe exponentielle. Les spins, initialement mis en phase par l'onde électromagnétique, vont progressivement se déphaser à cause des hétérogénéités locales du champ magnétique (relaxation spins-spins). De façon analogue à la séquence précédente, le temps T2 correspond au temps mis par l'aimantation transversale pour décroître de 37%. Moins il y a d'interactions entre particules, plus ce temps est long : le T2 est donc élevé dans les liquides (exemple : 3000 ms pour l'eau) et plus faible pour les tissus (exemple : 100 ms pour la substance blanche).

Il y a plusieurs réglages possibles pour l'acquisition de la séquence T2 (écho de spin, écho de gradient). La séquence T2 en écho de gradient (figure I.6, image du milieu) est préférée en raison d'une acquisition plus rapide. Afin de mieux identifier le contraste entre les tissus, le signal de l'eau peut être annulé à l'aide d'un réglage préliminaire donnant ainsi le paramètre T2 FLAIR (Fluid Attenuated Inversion Recovery). Cette séquence laisse apparaître en hypersignal les lésions cérébrales causées par l'AVC puisque l'œdème vasogénique se traduit par une augmentation du liquide intercellulaire, et donc un allongement du temps T2. Elle n'est cependant pas ou peu sensible à l'œdème cytotoxique parce que ce dernier entraîne des modifications de composition du milieu intercellulaire moins marquées. Ainsi, le T2 FLAIR est principalement utilisé pour le diagnostic final. A noter qu'il est aussi sensible à d'autres maladies notamment la maladie de la substance blanche dont le signal en T2 FLAIR est similaire aux lésions engendrées par l'AVC (figure I.7, troisième image).

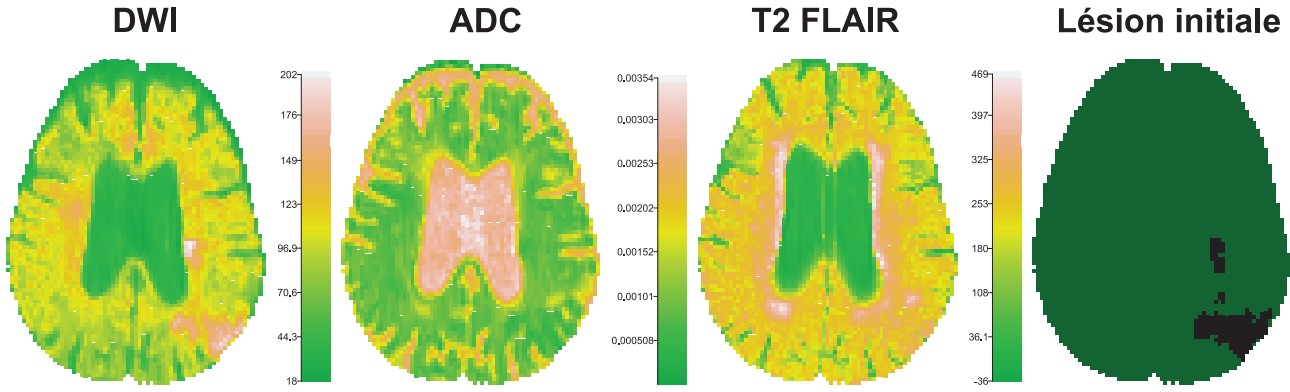


FIGURE I.7 – Paramètres caractérisant la lésion : séquence en diffusion (DWI à gauche et ADC au milieu gauche), séquence pondérée T2 où le signal du liquide a été annulé (T2 FLAIR au milieu droit) et la lésion finale (en noir à droite).

3.4 La séquence en diffusion

La séquence en diffusion permet d'obtenir un contraste très sensible aux variations de mobilité des molécules d'eau. Le principe est d'appliquer deux gradients symétriques dans une même direction : un premier déphase les particules proportionnellement à leur position dans l'axe du gradient puis, après un laps de temps, un second qui va rephaser les particules. Ainsi, si la particule ne s'est pas déplacée selon l'axe du gradient, elle est correctement rephasée. Dans le cas contraire le rephasage n'est pas correct et la résultante des moments perdra en intensité. Cette méthode est appliquée dans les 3 directions de l'espace afin de calculer la magnitude du déplacement. Cela permet d'en déduire un degré de mobilité : le coefficient de diffusion apparent (ADC, en $mm^2.s^{-1}$). En pratique un autre paramètre lui est préféré, le DWI, qui indiquera par des valeurs élevées les zones de lésion. Ces deux paramètres sont reliés par la formule (SCHAEFER et al., 2000) :

$$DWI = S_0 \exp(-b \text{ ADC})$$

où b est un paramètre d'acquisition reflétant l'intensité des gradients de diffusion et S_0 une constante déterminée à l'aide d'une acquisition supplémentaire (typiquement où $b = 0$).

À noter que le signal étant obtenu en appliquant des gradients à une séquence pondérée en T2, le contraste en DWI provient à la fois du contraste en diffusion (lié à la variation d'ADC) et du contraste en T2 (lié à l'intensité S_0). Aussi un hypersignal peut être dû à une restriction de diffusion mais aussi à une lésion déjà en hypersignal en T2. Cette séquence est donc utilisée pour détecter l'œdème cytotoxique causé par l'AVC en phase aiguë (figure I.7) puisque l'œdème provoque une restriction de mobilité des particules et donc une chute de l'ADC. Malgré son contraste T2, la séquence en diffusion perd en sensibilité avec le temps et, au-delà de quelques jours, la séquence T2 FLAIR lui est préférée.

3.5 La séquence en perfusion

L'objectif de la séquence en perfusion est de fournir des indicateurs de l'état de perfusion des tissus. L'idée générale est de marquer une partie du liquide sanguin et de caractériser son passage dans le volume considéré. Le marquage se fait ici à l'aide d'un produit de contraste paramagnétique, le gadolinium, qui va déphaser les spins et entraîner une diminution du temps de relaxation T2 des tissus environnants. Cette diminution permet de détecter le passage du liquide sanguin marqué dans le volume d'intérêt, le T2 étant généralement supposé varier linéairement avec la concentration en produit de contraste (ROSEN et al., 1990).

Toutefois les contraintes pratiques, notamment au niveau de l'injection, entraînent une légère dispersion du produit de contraste dans le sang. Ainsi plutôt que d'observer un instant précis d'arrivée, c'est une courbe d'évolution temporelle de la concentration en produit de contraste dans le volume d'intérêt qui est observée (figure I.9). Cette courbe pouvant être particulièrement bruitée, deux approches correctives ont été proposées. Une première approche consiste à ajuster une fonction gamma : $f(t) = A(t-B)^C \exp(-(t-D)/C)$ pour éliminer le bruit de mesure. Une seconde approche va prendre en compte la dispersion du produit de contraste à l'injection pour corriger la courbe de concentration. Pour cela, la courbe de concentration est déconvoluée par une fonction d'entrée artérielle (WU et al., 2003). Cette fonction d'entrée artérielle correspond à la dispersion de l'agent de contraste que l'on mesure dans une zone saine du cerveau, typiquement la zone controlatérale. A partir de ces méthodes, plusieurs valeurs résumées de la courbe sont extraites :

- **le volume sanguin cérébral (CBV)** correspondant au volume de liquide sanguin au sein du volume cérébral. Il est calculé par intégration numérique de la courbe de concentration (brute) du traceur.
- **le temps de pic (TTP)** correspondant au temps auquel le maximum de la concentration en produit de contraste est atteint. Il est obtenu en cherchant le maximum de la courbe de concentration modélisée à l'aide d'une fonction gamma.
- **le temps de maximum (TMAX)** est l'analogue du TTP sur la courbe de concentration déconvoluée.
- **le débit sanguin cérébral (CBF)** correspondant au débit sanguin cérébral au sein du volume cérébral. Il est calculé à partir de la courbe de concentration déconvoluée comme le débit maximal en produit de contraste (WU et al., 2003).
- **le temps de transit moyen (MTT)** correspondant au temps mis par le produit de contraste pour traverser le volume d'intérêt. Il est calculé comme le rapport entre le CBV et le CBF.

La figure I.8 représente les valeurs prises par chacun des paramètres de perfusion sur une coupe d'un patient. Les valeurs de perfusion sont sujettes à de nombreuses sources d'erreurs : non-respect de l'hypothèse de linéarité, dispersion du produit de contraste, effet de volume

partiel ; elle fait donc toujours l'objet de recherches importantes (voir l'article de JAHNG et al., 2014, pour une revue plus complète de la mesure de la perfusion en imagerie médicale).

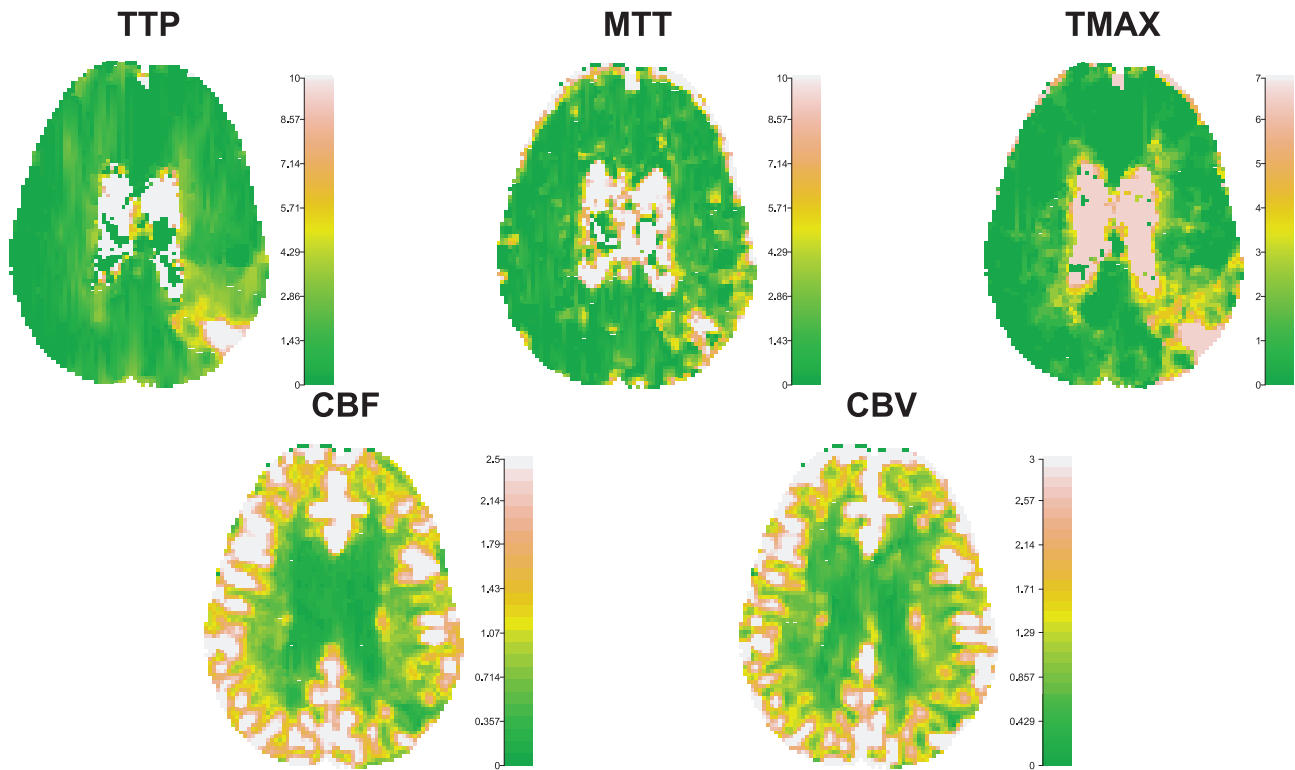


FIGURE I.8 – Paramètres de perfusion. De gauche à droite : temps jusqu'au pic de produit de contraste (TTP), temps de transit moyen (MTT), temps jusqu'au pic de produit de contraste après déconvolution (TMAX), débit sanguin cérébral (CBF) et volume sanguin cérébral (CBV).

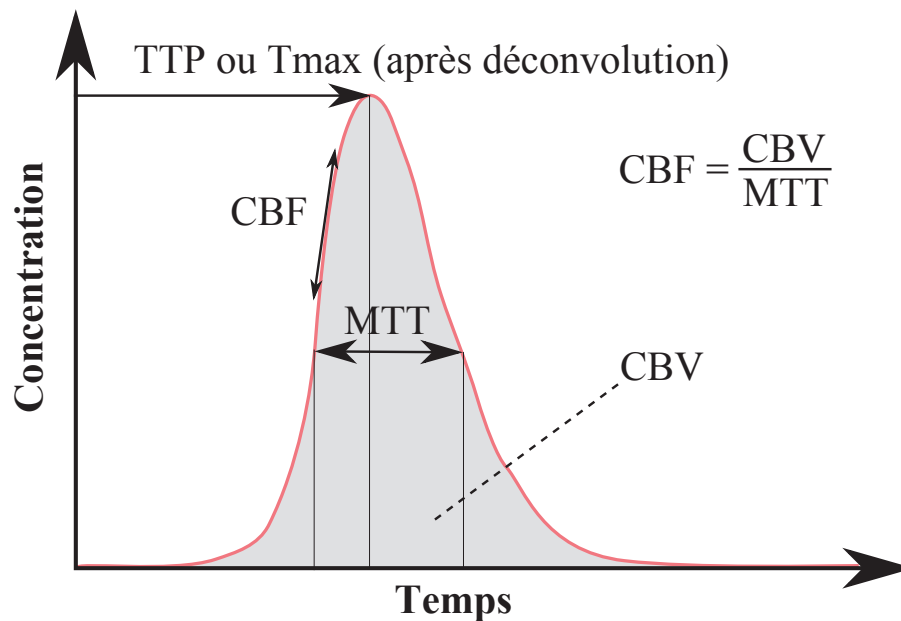


FIGURE I.9 – Courbe de concentration du produit de contraste dans le volume d'intérêt au cours du temps et lien avec les différents paramètres de perfusion.

4 Éligibilité des patients à la thrombolyse

Les critères d'éligibilité à la thrombolyse reflètent la balance bénéfice risque du traitement. En effet, un traitement trop tardif sera peu efficace et risque d'engendrer une hémorragie cérébrale (figure I.10). Initialement limitée à 3h après le début des symptômes, la thrombolyse est désormais recommandée jusqu'à 4h30 après le début des symptômes suite au résultat positif trouvé par l'étude ECASS III (LEES et al., 2010). Ainsi les principaux critères d'administration sont les suivants :

- Diagnostic d'un AVC confirmé par imagerie médicale (scanner ou IRM)
- Absence d'hémorragie cérébrale
- Délai depuis le début des symptômes connu et inférieur à 4h30

À ces critères s'ajoutent de nombreux critères d'exclusion (déficit neurologique mineur ou en régression, âge supérieur à 80 ans, déficit neurologique sévère, chirurgie majeure récente, ...).

Cependant, ces critères ne tiennent pas compte des spécificités de l'AVC (vitesse de croissance, taille de la zone de pénombre) et excluent d'office les patients endormis au moment de l'AVC (environ 20% des patients - WU et al., 2011). Ainsi des critères individualisés ont été proposés pour améliorer l'accès au traitement par les études DEFUSE (ALBERS et al., 2006) et DEFUSE 2 (LANSBERG et al., 2012). Trois profils ont été définis (figure I.11) : un profil malin, un profil mismatch et un profil non-mismatch. **Le profil malin** correspond aux AVC massifs qui sont à fort risque hémorragique et pour lesquels la thrombolyse sera insuffisante pour induire un bénéfice significatif. **Le profil mismatch** correspond aux patients ayant une zone de pénombre étendue et une lésion en diffusion modeste : ce sont les plus à même de répondre au traitement. Au contraire, **le profil non-mismatch** correspond à des patients ayant une petite zone de pénombre et qui ne retireront probablement aucun bénéfice de la thrombolyse. Cependant les critères d'imagerie caractérisant ces profils varient d'une étude à l'autre. Ainsi dans l'étude d'ALBERS et al., 2006, le profil mismatch est défini comme une lésion en diffusion entre 10 et 100 mL avec une zone de pénombre entre 10 et 100 mL dont le volume excède d'au moins 20% celui de la lésion en diffusion. Dans l'étude de LANSBERG et al., 2012, ce profil est défini comme une lésion de moins de 70 mL avec une zone de pénombre entre 15 et 100 mL dont le volume excède d'au moins 80% le volume de la lésion. Des critères plus consensuels pourraient être obtenus si ces derniers étaient définis par rapport au bénéfice minimum que doit apporter le traitement pour être cliniquement pertinent. Ce type de critère combiné à un modèle prédictif de l'évolution de l'AVC apparaît être une méthode prometteuse pour guider les indications thérapeutiques : une fois le bénéfice minimum fixé (par exemple 5 mL ou 25% de la lésion initiale), un patient serait traité si la différence d'évolution prédite par le modèle suite au traitement atteint ou dépasse le minimum fixé. Néanmoins cela suppose de disposer de modèles prédictifs fiables.

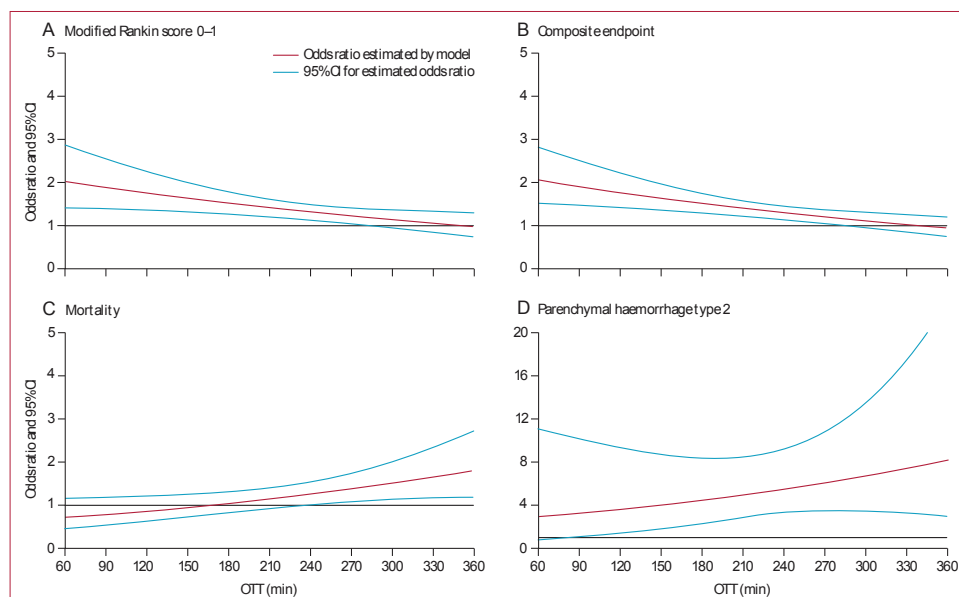


FIGURE I.10 – Impact du délai depuis les symptômes sur l'effet du traitement (LEES et al., 2010). Rapport des chances (ou "Odd ratio") de (A) réduction du handicap mesuré par le mRS, (B) réduction du handicap mesuré selon une échelle de score composite, (C) mortalité, (D) survenue d'une hémorragie cérébrale.

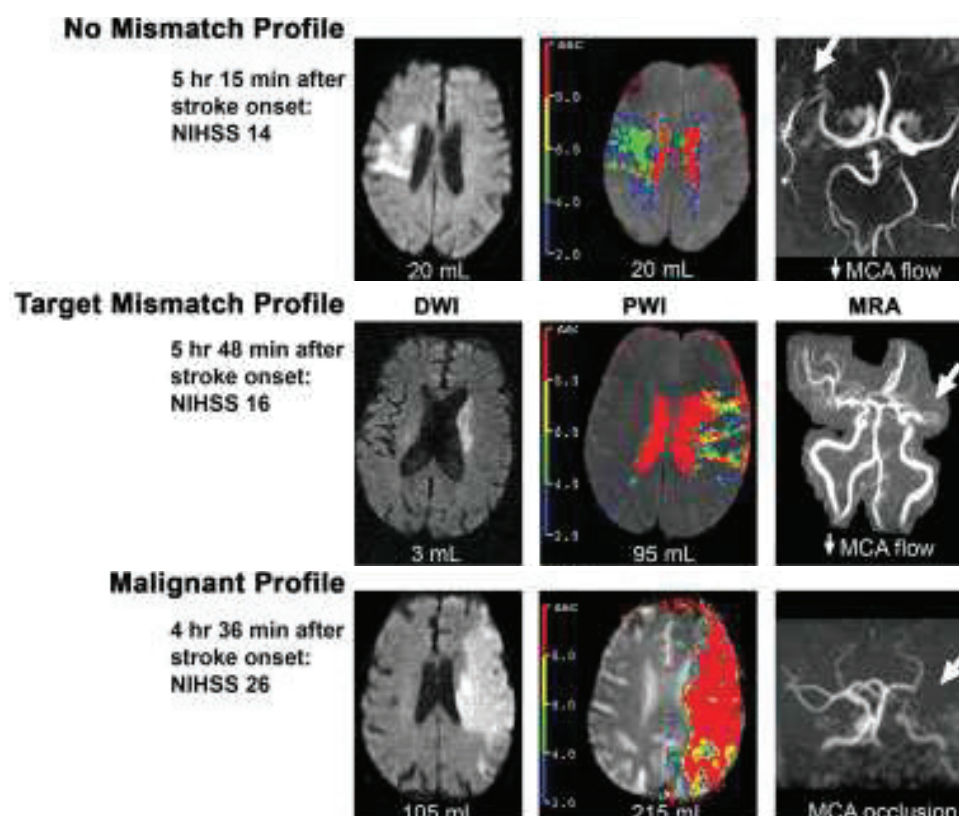


FIGURE I.11 – 3 profils proposés dans l'étude DEFUSE (ALBERS et al., 2006) :

- *Profil sans mismatch* où la lésion en diffusion (1^{er} ligne et 1^{er} colonne, en blanc) coïncide avec la lésion en perfusion (1^{er} ligne et 2nd colonne, en vert)
- *Profil avec mismatch* où la lésion en diffusion est entourée par la lésion en perfusion (2nd ligne et 2nd colonne, en jaune vert et bleu)
- *Profil malin* où l'AVC est très massif avec une lésion en diffusion occupant près de la moitié de l'hémisphère et une lésion en perfusion occupant quasiment tout l'hémisphère.

Partie II

Construction et évaluation de modèles prédictifs

1 Généralités sur la modélisation prédictive

1.1 Définition du modèle et de son ajustement

La modélisation prédictive a pour objectif de proposer une représentation simplifiée du mécanisme sous-jacent à une évolution. Cette représentation doit permettre de déduire, à partir d'un état initial, l'état final. Si l'on note \mathcal{X} le vecteur de variables aléatoires caractérisant l'état initial et Y l'état final, alors un modèle prédictif cherche à identifier une fonction de lien η entre \mathcal{X} et Y :

$$\text{chercher } \eta \text{ tel que : } Y = \eta(\mathcal{X})$$

Dans le cadre de l'AVC, les v.a. \mathcal{X} sont généralement des paramètres IRM ; dans le cas de variables cliniques elles seront notées \mathcal{C} . Par la suite on désignera par X_i la variable aléatoire correspondant au paramètre IRM X pour le voxel i et par $X = (X_i)_{i \in \{1, \dots, n\}}$ le vecteur des variables aléatoires associées aux différents voxels d'un patient pour le paramètre X . La variable Y sera le devenir du tissu contenu dans chaque voxel. Y est donc une variable binaire : le tissu contenu dans le voxel peut soit être sain ($Y = 0$) ou bien être nécrosé ($Y = 1$). Les problèmes de volume partiel, c'est-à-dire le fait qu'un voxel peut contenir différents types de tissu, ne seront pas considérés.

Le problème de prédiction se ramène donc à un problème de classification où l'on souhaite décider si Y prendra la valeur 0 ou 1. Pour cela, deux approches sont possibles pour construire la fonction η : soit directement prédire la valeur de Y , ou bien évaluer sur une échelle intermédiaire la propension qu'a Y à être égal à l'une des valeurs (typiquement 1). Dans ce second cas, on reformule le problème en introduisant une variable latente, Y^* , qui va servir d'échelle intermédiaire.

La classification sera ensuite effectuée en utilisant une règle de décision, typiquement en fixant un seuil τ à partir duquel on considère que la propension à appartenir à Y est suffisamment forte :

$$\begin{aligned} \text{Chercher } \eta \text{ et } \tau \text{ tel que : } Y^* &= \eta(\mathcal{X}) \\ \text{et } Y^* &> \tau = Y \end{aligned}$$

En statistique, la variable latente Y^* sera typiquement $\mathbb{P}[Y = 1 | \mathcal{X} = x]$, la probabilité que la v.a. Y prenne la valeur 1 connaissant la réalisation des v.a. \mathcal{X} . Mais d'autres variables latentes sont possibles telle qu'une distance signée au groupe de lésion. Cette seconde approche généralise la première : il est toujours possible à partir d'une variable latente de définir une règle de décision. Elle sera donc préférée par la suite. À noter que la plupart des méthodes de classification construisent une variable latente (même si elle n'est pas toujours visible par l'utilisateur) puis la dichotomisent avec un seuil fixé au préalable. Cette dichotomie peut se révéler sous-optimale lorsqu'elle ne prend en compte ni la prévalence des groupes, ni la structure de coût des erreurs.

En pratique, on ne peut pas espérer trouver une fonction η et un seuil τ vérifiant les égalités précédentes. La mesure IRM est en effet soumise à un certain nombre de sources de variabilité. Certaines sont structurelles : émission parasite de signal due à l'agitation thermique, différence de susceptibilité magnétique entre deux tissus entraînant des pertes de signal, modification de la fréquence de Larmor causée par les interactions avec les particules environnantes. D'autres sont pratiques : mauvaise uniformité du champ magnétique, mouvement du patient, dispersion du produit de contraste causé par le temps d'injection. Enfin certaines proviennent des étapes de traitement du signal : artéfacts de troncature de Gibbs, approximations faites par les méthodes de traitement du signal pour extraire les paramètres IRM. Plutôt qu'une résolution exacte, on va chercher à minimiser l'écart entre la grandeur observée et la grandeur prédite par le modèle :

$$\begin{aligned} \text{Étant donné } d(A, B) \text{ une mesure de l'écart entre deux v.a. } A \text{ et } B \\ \text{et } Y^* &= \eta(\mathcal{X}) \\ \text{chercher } \eta \text{ minimisant } d(Y^*, Y) \end{aligned}$$

On utilise alors le terme de **fonction objectif** pour désigner une fonction que l'on va maximiser ou minimiser afin d'identifier la fonction de lien. Ainsi en statistique, on cherche généralement à maximiser la vraisemblance des observations. L'utilisation d'une vraisemblance permet de s'adapter à un grand nombre de situations (données continues, binaires, de comptage, de temps jusqu'à événement, etc), ce qui fait sa popularité. La vraisemblance, notée L_v , est la probabilité que l'on a d'observer notre échantillon de valeurs selon notre modèle : elle est donc une mesure de compatibilité entre le modèle et l'échantillon. On peut alors se ramener au programme d'optimisation précédent en utilisant la déviance, $-2 * \log(L_v)$, comme fonction à minimiser. Le tableau II.1 page 17 indique pour les différents modèles prédictifs qui seront utilisés dans cette thèse les fonctions objectifs ainsi que les variables latentes associées. Les méthodes d'apprentissage automatique ont recours à des

quantités très hétérogènes tandis que les approches statistiques sont toutes basées sur l'optimisation d'une vraisemblance, éventuellement pénalisée.

L'espace des fonctions de lien η étant très vaste, il est généralement impossible d'effectuer le programme d'optimisation. Le modélisateur va donc se ramener à une famille de fonctions de lien dépendant d'un ou plusieurs paramètres notés θ :

Étant donné $d(A, B)$ une mesure de l'écart entre deux v.a. A et B
 et η une fonction de lien vérifiant $Y^* = \eta(\mathcal{X}, \theta)$
 chercher θ minimisant $d(Y^*, Y)$

En statistique, le choix du modèle de régression va définir le type distribution (et donc de fonction de lien). Le problème d'estimation des paramètres de la distribution peut alors se réécrire :

Étant donné une distribution \mathcal{L} telle que $Y \sim \mathcal{L}(\mathcal{X}, \theta)$ où θ est inconnu
 avec une fonction de lien η telle que $Y^* = \eta_\theta(\mathcal{X})$
 chercher θ maximisant $L_v(\theta|x, y) = \mathbb{P}[Y = y|\theta, \mathcal{X} = x]$

où x et y sont les réalisations des variables aléatoires \mathcal{X} et Y .

1.2 Prédiction pour une nouvelle observation

Une fois le modèle estimé, c'est-à-dire lorsque l'on a déterminé la valeur $\hat{\theta}$ de θ optimale pour les réalisations x et y considérées, il devient simple de formuler une prédiction pour une nouvelle observation. Pour cela, il suffit de calculer la valeur de la variable latente en utilisant la variable x_0 mesurée pour la nouvelle observation : $y_0^* = \eta(x_0, \hat{\theta})$. Si l'on désire une classification "dure", on pourra alors appliquer la règle de décision $y_0^* > \tau$ pour déterminer la classe de la nouvelle observation. Cependant nous nous limiterons (sauf mention contraire) à une classification "floue" en considérant la v.a. Y^* comme **un biomarqueur fictif du risque de nécrose**. Le modèle prédictif ainsi obtenu permet d'agréger l'information relative au patient en un biomarqueur du risque de nécrose de façon optimale au sens de la fonction objectif utilisée. La valeur y_0^* correspond donc à une prédiction quantitative du risque de nécrose associé à chaque voxel.

1.3 Évaluation d'un modèle prédictif

L'ajustement d'un modèle prédictif ne préjuge en rien de sa qualité. L'évaluation du modèle est donc une étape importante et peut correspondre à deux objectifs : évaluation absolue des performances ou comparaison du modèle vis-à-vis d'un autre modèle. Dans le premier cas, on cherche à vérifier que Y^* est proche de Y . Dans l'autre cas, on cherche à savoir si le biomarqueur associé à un premier modèle $Y_{\mathcal{M}1}^*$ est plus proche de Y que le biomarqueur associé à un second modèle $Y_{\mathcal{M}2}^*$.

Modèle	Abréviation	Fonction objectif (d ou L_v)	Variable latente (Y^*)
Modèle logistique	LR	Vraisemblance	Probabilité de nécrose
Modèle logistique à effets mixtes	MLR	Vraisemblance pénalisée par l'écart aux effets moyens	Probabilité de nécrose
Modèle logistique additif	ALR	Vraisemblance pénalisée par la régularité des fonctions de risque	Probabilité de nécrose
Algorithme Adaboost	ADA	Erreurs de classification	Vote pondéré des arbres
Réseau de neurones artificiels	ANN	Erreurs de classification	Somme pondérée des entrées
Forêts d'arbres décisionnels	RF	Indice de Gini	Vote des arbres
Machines à vecteurs de supports	SVM	Marge entre les classes, pénalisée par les erreurs de classification	Distance à un hyperplan

TABLE II.1 – Modèles prédictifs utilisés dans la section 3. Les modèles de régression et l'algorithme Adaboost sont présentés en annexe B.

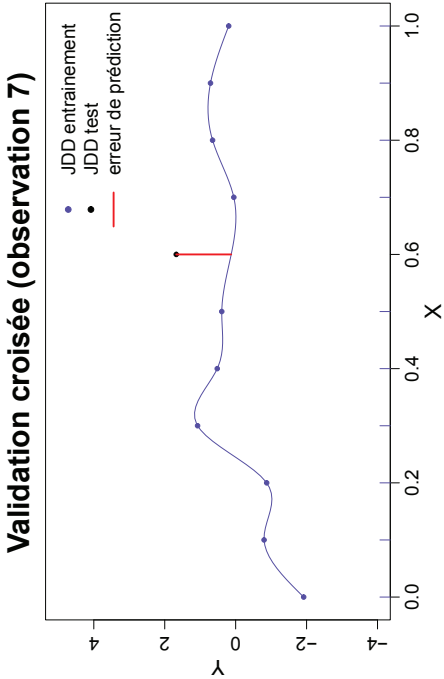
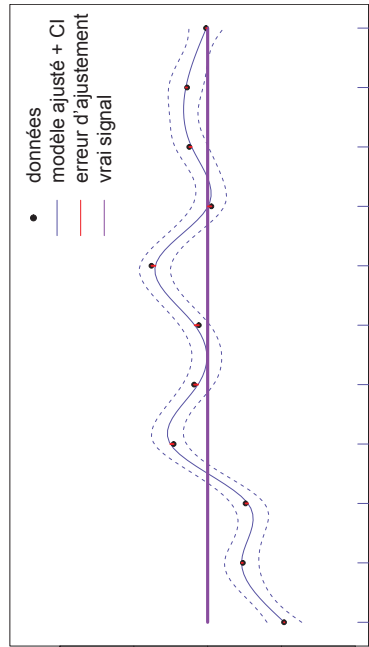


FIGURE II.1 – À gauche exemple de surajustement : ajustement d'une fonction souple (en bleu, traits pleins) et son intervalle de confiance (en bleu, traits pointillés) sur un échantillon de données (en noir) simulé à partir d'une fonction théorique (en violet) en ajoutant des erreurs corrélées dans le temps. À droite : erreur de prédiction après avoir retiré la 7^{ème} donnée du jeu de données d'entraînement.

Il reste donc à définir une mesure de proximité entre Y et Y^* . Une première approche consisterait à utiliser les valeurs des fonctions objectifs après optimisation mais, dans un cadre prédictif, elle a de plusieurs limites. D'une part, elle évalue l'ajustement du modèle et non sa capacité prédictive. Or, comme le montre la figure II.1 (graphique de gauche), un modèle peut être bien ajusté aux données tout en ayant une mauvaise capacité prédictive : on parle alors de **surapprentissage**. D'autre part, la fonction objectif est généralement peu interprétable en tant que telle : une déviance de 307425 est peu informative sur la qualité du modèle. Plus gênant, les fonctions objectifs ne sont souvent pas comparables entre elles : comparer une déviance de 307425 avec 91,5% d'observations bien classées fait peu de sens. Une exception notable concerne les modèles de régression dit "emboîtés", c'est-à-dire lorsqu'un modèle correspond à un cas particulier d'un autre modèle et que tous deux sont estimés en maximisant la vraisemblance. Dans ce cas il est possible de comparer l'ajustement entre les deux modèles à l'aide d'un test du ratio de vraisemblance.

Ainsi, il est nécessaire d'utiliser d'autres métriques pour évaluer et comparer la qualité prédictive d'un modèle. Pour cela, différents indices ont été définis, parfois plus spécifiquement pour un domaine d'application. L'épidémiologie utilise préférentiellement la sensibilité (Se), la spécificité (Sp), la valeur prédictive positive (PPV), la valeur prédictive négative (NPV) et l'aire sous la courbe ROC (AUC) pour évaluer ces modèles. En imagerie, l'évaluation se fait plus volontiers selon des indices de recouvrement volumique tel que l'indice de Dice ou l'indice de Jaccard. La définition et le choix des indices pertinents dans le cadre de modèles prédictifs pour l'AVC seront discutés dans la section suivante (section 2).

1.4 Optimisme et correction par rééchantillonnage

Comme l'illustre la figure II.1, un modèle ne doit pas être évalué sur le jeu de données qui a été utilisé pour l'ajuster. En effet, les données étant bruitées, le modèle va s'ajuster sur le signal mais aussi sur le bruit. L'ajustement sera donc particulièrement performant sur ce jeu de données mais peut être mauvais sur un autre jeu de données puisque, par définition, le bruit varie d'un jeu de données à l'autre. En évaluant le modèle sur le jeu de données d'entraînement on surévaluerait les performances du modèle : cela s'appelle l'optimisme.

L'idéal serait donc d'évaluer le modèle sur un jeu de données externe. Un tel jeu de données étant rarement disponible, différentes techniques de rééchantillonnage ont été proposées pour générer artificiellement un jeu de données externe : partitionnement des données, validation croisée, rééchantillonnage (voir l'article de MOLINARO et al., 2005 pour plus de détails). Le partitionnement de données, technique consistant à diviser le jeu de données en un jeu de données d'entraînement et un jeu de données test, aboutit généralement à une moins bonne estimation des performances du modèle comparé à la validation croisée ou au rééchantillonnage. Ces deux dernières méthodes sont considérées de qualité équivalente, le compromis biais/variance de l'estimateur des performances dépendant du paramétrage utilisé pour chaque méthode. **La validation croisée** partitionne les données en un nombre ν de

partitions, ν étant fixé à l'avance. Chaque partition est alors utilisée successivement comme jeu de données test, le reste étant utilisé pour ajuster le modèle (figure II.1, à droite). Le rééchantillonnage suit la même approche sauf qu'au lieu de partitionner le jeu de données initial, il va former des groupes de données par tirage aléatoire dans les données de départ.

2 L'évaluation des modèles prédictifs pour l'AVC

Dans cette section, nous supposons disposer des valeurs de risque de nécrose pour chaque voxel de chaque patient. Ces valeurs forment un biomarqueur, fictif s'il est issu du modèle de prédiction ou réel s'il correspond à un paramètre de perfusion, dont nous souhaitons évaluer la capacité prédictive de l'AVC final.

2.1 Prise en compte de la structure des données dans l'évaluation

Avec la résolution spatiale de nos données, nous disposons de l'ordre de 30 000 voxels par patient dans chaque hémisphère et donc approximativement 1 300 000 voxels pour l'ensemble des patients. Si l'on considère disposer de $n=1\,300\,000$ observations indépendantes, l'incertitude sur les estimations statistiques sera infime puisque celle-ci évolue généralement en $n^{-1/2}$.

Or cette approche néglige complètement la variabilité inter-patients qui, comme nous le verrons par la suite, est en réalité importante. Nos données ont bien deux niveaux hiérarchiques : un niveau patient et un niveau voxel (figure II.2). Les données au niveau voxel sont corrélées au sein de chaque patient : elles ne sont donc pas indépendantes et le nombre d'observations surestime le nombre de degrés de liberté. Ainsi pour avoir une information complète il faudrait à la fois avoir une résolution IRM parfaite (faire tendre le nombre de voxels par patient vers l'infini) et avoir un nombre infini de patients.

Nous choisirons de considérer le niveau patient comme unité statistique de référence pour l'évaluation des modèles. En mesurant les performances du modèle chez chaque patient et en supposant les patients indépendants, nous obtiendrons des mesures indépendantes de la performance de notre modèle. Il sera alors possible de comparer les performances de deux modèles comme la comparaison de deux échantillons appariés et de décrire les performances d'un modèle à l'aide des statistiques descriptives usuelles pour un échantillon de données.

Cette approche ne prend en compte que l'incertitude inter-patients et néglige l'incertitude intra-patient. Cette approximation est motivée par le fait que le très grand nombre de voxels par patient rend l'estimation intra-patient précise, laissant l'incertitude inter-patients prépondérante. Néanmoins, il s'agit bien d'une approximation puisque la vraie incertitude devrait bien être une somme des incertitudes inter-patients et intra-patient.

2.2 Indices de performance des modèles prédictifs

Le tableau II.2 présente les critères utilisés dans la littérature pour évaluer les approches prédictives voxel à voxel. Ce tableau n'est pas exhaustif : lorsqu'une même équipe a publié successivement plusieurs travaux, seul le plus récent est reporté. De plus, seules les études portant sur des modèles prédictifs voxel à voxel sont reportées.

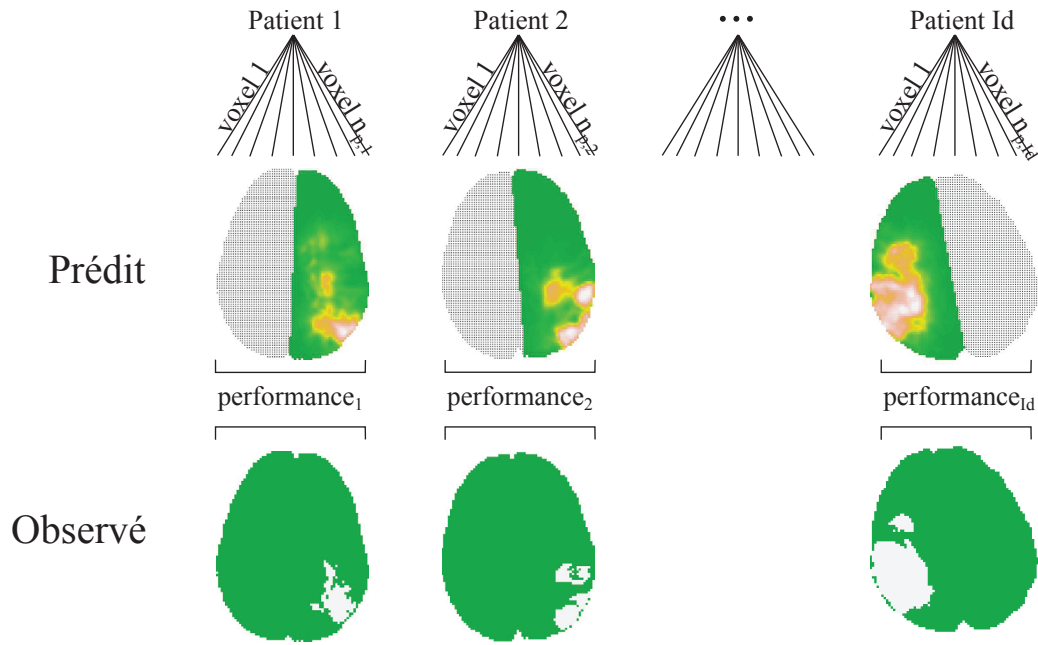


FIGURE II.2 – Représentation schématique de l'évaluation des performances du modèle prédictif. Pour chaque patient, les valeurs de biomarqueur de risque prédit pour les voxels du patient sont comparées à une référence.

Article	Indice(s) principal(aux)	Indice(s) secondaire(s)
CARANO et al., 2000	%BC	$\rho(V, \hat{V})$
LU et al., 2005	$\rho(V, \hat{V})$	
SHEN et al., 2008a	AUC, Se, Sp	
ROSSO et al., 2009	$\rho(V, \hat{V})$	
SCALZO et al., 2012	AUC	
BAGHER-EBADIAN et al., 2011	AUC	%BC, Se, Sp, corrélation prédictions et T2
BOUTS et al., 2013	AUC, AUPRC, Dice	Se, Sp
GOTTRUP et al., 2005	AUC	
HUANG et al., 2011	AUC	
KIDWELL et al., 2013	%BC	AUC
NGUYEN et al., 2008	AUC, %BC	
WU et al., 2001	AUC, Se, Sp	
WU et al., 2004	AUC, %BC	
WU et al., 2006	AUC	Se, Sp, $ \hat{V} - V $

TABLE II.2 – Indices d'évaluation des approches prédictives de la nécrose des tissus utilisés dans la littérature. Les principales caractéristiques des approches prédictives correspondantes sont reportées dans le tableau II.4 page 44.

Au vu du tableau, il apparaît clairement que l'AUC est l'indice privilégié. Néanmoins, d'autres indices sont souvent utilisés en complément de l'AUC. Leur hétérogénéité reflète bien que les approches prédictives proposées (qui seront présentées ultérieurement, voir tableau II.4 page 44) sont à la croisée de plusieurs domaines : les statistiques avec des modèles de régression et de classification, la recherche d'information avec des méthodes d'apprentissage automatique, l'imagerie avec des méthodes de traitement d'images et de reconnaissance de formes. Ce foisonnement d'indices pouvant être source de confusion, ils sont brièvement présentés dans les trois paragraphes suivants.

Indices utilisés pour l'évaluation de biomarqueurs

Les indices issus de la courbe ROC sont couramment utilisés en santé pour évaluer des biomarqueurs pronostiques ou diagnostiques. Ces indices étant très classiques nous ne les exposerons ici que brièvement, le lecteur intéressé pourra trouver une introduction plus complète dans l'article de FAWCETT, 2006.

Considérons une variable Y à deux états et un biomarqueur Y^* dont on souhaite évaluer la capacité à discriminer les voxels sains ($Y = 0$) des voxels malades ($Y = 1$). Pour un seuil de biomarqueur τ , on peut former la table de contingence associée (tableau II.3). Grâce à cette table, il est possible de calculer différents indices qui vont nous renseigner sur la capacité discriminante du biomarqueur considéré :

- **la sensibilité (Se)** qui évalue la capacité du test à repérer les voxels malades :

$$Se = \mathbb{P}[Y^* \geq \tau | Y = 1] = \frac{TP}{TP + FN}$$

- **la spécificité (Sp)** qui évalue la capacité du test à repérer les voxels sains :

$$Se = \mathbb{P}[Y^* < \tau | Y = 0] = \frac{TN}{TN + FP}$$

- **la valeur prédictive positive (PPV)** qui indique la confiance que l'on peut avoir dans un test positif :

$$PPV = \mathbb{P}[Y = 1 | Y^* \geq \tau] = \frac{TP}{TP + FP}$$

- **la valeur prédictive négative (NPV)** qui indique la confiance que l'on peut avoir dans un test négatif :

$$NPV = \mathbb{P}[Y = 0 | Y^* < \tau] = \frac{TN}{TN + FN}$$

- le **pourcentage de voxels bien classés** (%BC) qui indique la qualité de la classification binaire :

$$\%BC = \mathbb{P}[Y = (Y^* \geq \tau)] = \frac{TP + TN}{TP + FP + FN + TN}$$

- l'**aire sous la courbe ROC (AUC)** qui correspond à la probabilité que les voxels sains aient une valeur de biomarqueur plus faible que les voxels malades. Contrairement aux indices précédents qui étaient calculés pour un seuil τ fixé, l'AUC résume la capacité discriminante du biomarqueur sur l'ensemble des seuils possibles :

$$AUC = \int_{\tau=+\infty}^{-\infty} Se(\tau) d\{1 - Sp(\tau)\}$$

L'intégration se fait relativement aux accroissements de sensibilité et donc, depuis les seuils élevés vers les seuils faibles.

		États observés	
		Malade ($Y = 1$)	Sain ($Y = 0$)
Valeurs de biomarqueur	Positif ($Y^* \geq \tau$)	Vrai Positif (TP) 759	Faux Positif (FP) 364
	Négatif ($Y^* < \tau$)	Faux Négatif (FN) 656	Vrai Négatif (TN) 30 199

TABLE II.3 – Table de contingence. En noir, noms génériques et en gris, valeurs obtenues lors de l'évaluation du modèle logistique chez un patient avec un seuil τ de 0,42.

Dans cet exemple : Se=0,536 Sp=0,988 PPV=0,676 NPV=0,979
 %BC=0,968 Jaccard=0,427 AUC=0,964 AUPRC=0,642
 prévalence=0,044

Généralement, l'évaluation de biomarqueurs se fait en considérant le compromis sensibilité-spécificité, à un seuil donné ou en utilisant l'AUC. Ces grandeurs ont l'avantage d'être indépendantes de la prévalence puisqu'elles sont calculées au sein de chaque sous-population (malade ou sain) : leurs valeurs sont supposées refléter les caractéristiques intrinsèques du test et non être spécifiques à la population étudiée. Cette pratique est toutefois questionnée, notamment dans le cadre de faibles prévalences où il est recommandé d'examiner aussi les valeurs prédictives positives et négatives (GRIMES et al., 2002).

Indices utilisés pour l'évaluation des méthodes de recherche d'information

Dans le domaine de la recherche de l'information, il y a généralement un faible nombre d'informations importantes et un grand nombre d'informations à éliminer (GOADRICH et al., 2004). Il s'agit donc d'évaluer la méthode selon sa capacité à identifier correctement l'information pertinente. Les vrais négatifs, souvent très nombreux, sont peu informatifs sur la qualité de la méthode.

Ainsi plutôt que d'étudier le compromis sensibilité-spécificité, l'évaluation se fait généralement au regard des indices précision et rappel (MANNING et al., 2008) :

- **la précision** est un autre nom donné à la VPP.
- **le rappel** est un autre nom donné à la sensibilité.
- **l'AUPRC** est l'analogue de l'AUC pour la courbe précision-rappel (PR) :

$$AUPRC = \int_{\tau=-\infty}^{-\infty} VPP(\tau) d\{Se(\tau)\}$$

Comme pour l'AUC, l'intégration se fait relativement aux accroissements de sensibilité.

Différents indices ont aussi été proposés pour traduire le compromis précision-rappel à seuil fixé (comme la F-mesure) mais ils ne seront pas considérés ici.

Indices utilisés pour l'évaluation des volumes en imagerie

Mesurer la capacité discriminante du biomarqueur permet de s'assurer de la qualité des risques de nécrose prédits et de la bonne correspondance spatiale avec la lésion finale. Néanmoins même si la visualisation des risques prédits sur les différentes coupes IRM du patient permet de représenter l'évolution de l'AVC, il ne s'agit pas d'un outil d'aide à la décision facile à utiliser en pratique clinique. Le volume final de l'AVC ou le volume d'expansion de l'AVC sont des notions plus concrètes et plus utilisables pour le médecin. Aussi le modèle prédictif doit être évalué au regard de sa précision volumique c'est-à-dire au regard de l'accord entre le volume d'AVC final et le volume d'AVC prédit. Pour cela, différentes mesures peuvent être considérées :

- **l'erreur volumique en valeur absolue**, notée $|\hat{V} - V|$, quantifie l'erreur (en valeur absolue) entre le vrai volume (V) et le volume prédit (\hat{V}).
- **l'indice de Jaccard** mesure le recouvrement entre deux ensembles, un ensemble de référence (A) et un ensemble à évaluer (B) :

$$|Jaccard| = \frac{A \cap B}{A \cup B} = \frac{TP}{TP + FP + FN}$$

L'indice de Dice, aussi très utilisé, capte essentiellement la même information que l'indice de Jaccard (BOUIX et al., 2007).

- **la corrélation entre les volumes prédits et observés**, notée $\rho(V, \hat{V})$.

2.3 Les limites des indices issus de la courbe ROC

Comme nous l'avons vu dans le tableau II.2, l'évaluation des modèles prédictifs au niveau voxel se fait quasi-systématiquement au regard de l'AUC, ou dans certains cas en considérant le pourcentage de bien classés. L'AUC est certes un indice largement étudié et utilisé mais son utilisation systématique est fréquemment remise en cause (DAVIS et al., 2006 ; VAN CALSTER et al., 2013).

Les AUC des modèles prédictifs pour l'AVC reportées dans la littérature sont très élevées (entre 0,8 et 0,9), conduisant parfois à des conclusions optimistes : "This result implies that the trained ANN can provide an estimate of 3-month ischemic lesion on T2WI in a stable and accurate manner (AUROC = 0,89)" (BAGHER-EBADIAN et al., 2011) ou "Predictions were overall highly accurate." (HUANG et al., 2010) ou encore "Voxel-based algorithms using acute multiparametric-MRI data have been shown to accurately predict tissue outcome after stroke." (WU et al., 2012). Cela contraste avec la précision volumique qui, lorsqu'elle était reportée, était très médiocre. Ainsi WU et al., 2006, reportent pour un groupe de patients non-traités une AUC de $0,9 \pm 0,05$ avec une erreur volumique de 37 ± 35 mL (pour des volumes finaux de 85 ± 88 mL) soit une erreur relative moyenne d'environ 40%. Une telle approche ne semble donc pas être en mesure de fournir des estimations fiables du volume de pénombre. Elle a néanmoins été proposée pour évaluer l'efficacité d'un traitement sur le principe suivant : le traitement modifiant l'évolution de l'AVC, une diminution des capacités prédictives du modèle chez les patients traités est imputable à un effet du traitement (WU et al., 2012). Néanmoins cela suppose au moins que le modèle prédictif initial soit correct et montre l'importance d'une évaluation convenable de ce dernier.

La mauvaise qualité des volumes prédits peut résulter de l'inexactitude des prédictions fournies par le modèle mais aussi de la méthode employée pour passer des probabilités de nécrose à un volume. Dans l'article de WU et al., 2006, le volume était calculé en comptant le nombre de voxels ayant un risque supérieur à 0,5. L'imprécision liée à cette méthode pourrait expliquer l'incohérence entre la précision volumique et l'AUC. Néanmoins, l'utilisation de l'AUC en imagerie médicale pose question. En effet, la zone imagée (ici le cerveau) est généralement bien plus importante que la zone touchée par la maladie : cela conduit à une faible proportion de voxels affectés par la maladie comparé au grand nombre de voxels sains. Une méthode pour être pertinente devra avoir une très bonne spécificité. Ainsi seule la zone de la courbe ROC correspondant aux fortes spécificités est intéressante alors qu'elle n'influe qu'assez faiblement dans le calcul de l'AUC. Dans un tel cas, plusieurs approches sont possibles pour mesurer plus fidèlement les performances :

- **Se placer à un seuil**, typiquement correspondant à une spécificité élevée, et étudier les indices de performance associés au seuil (PPV, NPV, Se, Sp). Le principal inconvénient de cette approche est qu'elle nécessite le choix d'un seuil, ce qui est délicat. Le seuil "optimal" peut changer suivant le modèle retenu rendant les comparaisons entre modèles difficiles. D'autre part, cela ne remplit pas l'objectif souhaité qui était d'évaluer la distribution de

risque Y^* et non une version dichotomisée $Y^* > \tau$. Cela peut se révéler encore plus gênant si le calcul du volume ne se fait pas par seuillage mais via une autre méthode (voir section 3.6 page 87).

- **Utiliser une aire sous la courbe partielle.** Il est possible de n'évaluer la courbe ROC du biomarqueur (et donc l'aire sous la courbe) que sur une portion de spécificité (WALTER, 2005), par exemple uniquement pour les spécificités élevées. Le principal inconvénient de cette approche est qu'elle nécessite de définir un intervalle de spécificité d'intérêt. Cela rend l'évaluation subjective au choix de l'intervalle.
- **Utiliser la courbe précision-rappel.** Cette approche est préconisée dans la littérature de recherche d'information lorsque que la prévalence est faible (DAVIS et al., 2006; FAWCETT, 2006; BOYD et al., 2013). Néanmoins excepté quelques propriétés théoriques (DAVIS et al., 2006; BOYD et al., 2013), le comportement de l'aire sous la courbe PR (AUPRC) est peu connu.

Cette dernière approche semble être la plus appropriée à l'évaluation objective des performances. Néanmoins avant de l'utiliser, il nous a semblé important de mieux comprendre le comportement de l'AUPRC et de le comparer à celui de l'AUC à l'aide d'une étude de simulation.

2.4 Comparaison de la pertinence de l'AUC et de l'AUPRC pour l'évaluation d'un biomarqueur

L'objectif de l'étude était de fournir des résultats assez généraux concernant le comportement de l'AUC et de l'AUPRC, en incluant le cas des faibles prévalences. Pour cela nous avons simulé des valeurs de biomarqueur pour un groupe sain et un groupe malade en considérant plusieurs prévalences et plusieurs paramétrisations de la distribution du biomarqueur dans chaque groupe. Puis nous avons étudié si les valeurs d'AUC et d'AUPRC reflétaient bien la qualité du biomarqueur. Les deux paragraphes suivants présentent les grandes lignes de l'étude ; pour davantage de détails le lecteur est invité à consulter l'article correspondant à la page 29.

Quelques considérations théoriques

L'AUC, et plus généralement l'approche ROC, peut être vue comme une mesure du recouvrement entre la distribution du biomarqueur chez les sujets sains et chez les sujets malades. En effet, l'AUC correspond à la probabilité qu'un sujet malade ait une valeur de biomarqueur plus élevée qu'un sujet sain (BAMBER, 1975) :

$$\begin{aligned}
 AUC &= \int_{\tau=-\infty}^{-\infty} Se(\tau) d\{1 - Sp(\tau)\} \\
 &= \int_{\tau=-\infty}^{-\infty} \mathbb{P}[Y^* \geq \tau | Y = 1] d\{1 - \mathbb{P}[Y^* < \tau | Y = 0]\} \\
 &= \int_{\tau=-\infty}^{-\infty} \mathbb{P}[Y_{Y=1}^* \geq \tau] d\{1 - \mathbb{P}[Y_{Y=0}^* < \tau]\} \\
 &= \int_{\tau=-\infty}^{+\infty} \mathbb{P}[Y_{Y=1}^* \geq \tau] dF_{Y_{Y=0}^*}(\tau) \\
 &= \mathbb{P}[Y_{Y=1}^* \geq \tau | \tau \sim Y_{Y=0}^*] \\
 &= \mathbb{P}[Y_{Y=1}^* > Y_{Y=0}^*]
 \end{aligned}$$

Une interprétation de l'AUPRC, certes plus complexe que celle de l'AUC, a été proposée par BOYD et al., 2013 : l'AUPRC correspond à l'espérance de la proportion de malades, parmi l'ensemble des sujets ayant une valeur de biomarqueur au-delà d'un seuil échantillonné parmi les valeurs de biomarqueur des malades :

$$\begin{aligned}
 AUPRC &= \int_{\tau=-\infty}^{-\infty} Precision d\{Recall\} \\
 &= \int_{\tau=-\infty}^{-\infty} VPP(\tau) d\{Se(\tau)\} \\
 &= \int_{\tau=-\infty}^{-\infty} \mathbb{P}[Y = 1 | Y^* \geq \tau] d\mathbb{P}[Y^* \geq \tau | Y = 1] \\
 &= - \int_{\tau=-\infty}^{-\infty} \mathbb{P}[Y = 1 | Y^* \geq \tau] dF_{Y_{Y=1}^*}(\tau) \\
 &= \int_{\tau=-\infty}^{+\infty} \mathbb{P}[Y = 1 | Y^* \geq \tau] dF_{Y_{Y=1}^*}(\tau) \\
 &= \mathbb{P}[Y = 1 | Y^* \geq \tau, \tau \sim Y_{Y=1}^*]
 \end{aligned}$$

puisque $dF_{Y_{Y=1}^*}(\tau) = d\mathbb{P}[Y_{Y=1}^* \leq \tau] = d\{1 - \mathbb{P}[Y_{Y=1}^* \geq \tau]\} = -d\mathbb{P}[Y_{Y=1}^* \geq \tau]$

Ainsi l'AUC est indépendante de la prévalence puisqu'elle compare les distributions. L'AUPRC regarde la proportion de malades sur les seuils "à risque" et est donc dépendante de la prévalence. L'AUPRC peut être vue comme une mesure du recouvrement de l'histogramme des sujets malades par celui des sujets sains.

Malgré ces différences, il y a des liens forts entre les approches ROC et PR (DAVIS et al., 2006) : à prévalence fixée les deux courbes contiennent la même quantité d'information. Par ailleurs, si les courbes ne se croisent pas, les AUC et AUPRC de deux biomarqueurs sont rangées dans le même sens.

Principaux résultats

Les principaux résultats de l'article se retrouvent sur la figure 2 de l'article (page 32) :

- en cas de prévalence proche de 0,5, l'AUC et l'AUPRC sont concordantes.
- en cas de faible prévalence, l'AUC peut donner une vision trop optimiste des performances. Par exemple, l'AUC peut être très élevée (0,962) et la PPV très faible ($<0,25$ sur presque tous les seuils). L'AUPRC semble correctement mesurer le compromis PPV, NPV même à de faibles prévalences.
- Le principal inconvénient de l'AUPRC est qu'étant dépendante de la prévalence, elle peut difficilement être utilisée pour comparer les performances issues de populations différentes (excepté si la prévalence est la même, ce qui est rare). Il reste donc toujours intéressant de reporter l'AUC pour permettre de comparer différentes études.

Article publié dans the Journal of Clinical Epidemiology

Partant du constat que la courbe PR est peu employée par la communauté médicale et donc probablement méconnue : une recherche PubMed réalisée en janvier 2015 avec les mots clés : "Precision" "Recall" et "Curve" ne donnait que 77 résultats et seulement 9 d'entre eux correspondaient à des travaux en santé, la publication de ce travail avait pour but de montrer les limites de l'AUC dans le cas de faibles prévalences mais aussi de rappeler l'existence d'outils alternatifs. Un format court a été privilégié pour assurer un message concis au lecteur. Les 5 pages qui suivent contiennent le corps de l'article. Une analyse supplémentaire avait été demandée par un des relecteurs du journal et a été mise en matériel supplémentaire.



The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases

Brice Ozenne^{a,b,c,d}, Fabien Subtil^{a,b,c,d}, Delphine Maucourt-Boulch^{a,b,c,d,*}

^aUniversité de Lyon, Département Biomaths-Santé, Lyon, 92 Rue Pasteur, 69007, France

^bUniversité Lyon 1, Département Biomaths-Santé, Villeurbanne, 43 boulevard du 11 Novembre 1918, 69622, France

^cHospices Civils de Lyon, Service de Biostatistique Lyon, 165 Chemin du Grand Revoyet, Pierre-Bénite F-69310, France

^dCNRS UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Equipe Biostatistique-Santé, Villeurbanne F-69100, France

Accepted 23 February 2015; Published online 28 February 2015

Abstract

Objectives: Compare the area under the receiver operating characteristic curve (AUC) vs. the area under the precision–recall curve (AUPRC) in summarizing the performance of a diagnostic biomarker according to the disease prevalence.

Study Design and Setting: A simulation study was performed considering different sizes of diseased and nondiseased groups. Values of a biomarker were sampled with various variances and differences in mean values between the two groups. The AUCs and the AUPRCs were examined regarding their agreement and vs. the positive predictive value (PPV) and the negative predictive value (NPV) of the biomarker.

Results: With a disease prevalence of 50%, the AUC and the AUPRC showed high correlations with the PPV and the NPV ($\rho > 0.95$). With a prevalence of 1%, small PPV and AUPRC values (< 0.2) but high AUC values (> 0.9) were found. The AUPRC reflected better than the AUC the discriminant ability of the biomarker; it had a higher correlation with the PPV ($\rho = 0.995$ vs. 0.724 ; $P < 0.001$).

Conclusion: In uncommon and rare diseases, the AUPRC should be preferred to the AUC because it summarizes better the performance of a biomarker. © 2015 Elsevier Inc. All rights reserved.

Keywords: Area under the curve; Binary biomarker; Performance assessment; Precision-Recall curve; Rare events; Receiver operating curve

1. Introduction

Assessing the performance of a diagnostic biomarker and comparing the performances of several biomarkers are important issues in medicine. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) are recommended indicators that evaluate diagnostic performance [1]. The PPV and the NPV are particularly interesting in case of low disease prevalence [2]; however, they rely on the choice of a threshold biomarker value that classifies the subjects into diseased and nondiseased. As the optimal threshold may depend on various characteristics of the population, global measures have been proposed.

The receiver operating characteristic (ROC) curve is often considered as the standard mean for depicting biomarker performance over all biomarker thresholds. Its properties and associated indices have been extensively

studied [3], but its relevance in case of rare events is still debated. Indeed, only a few works have dealt with rare diseases or medical imaging data whose prevalence is typically very low (eg, Wilson disease whose worldwide prevalence is estimated at 30 per 1 million [4] or the proportion of lesioned to nonlesioned pixels on stroke images that is frequently lower than 1% [5]).

In the field of information retrieval where there are important differences between the small number of relevant documents and the huge number of irrelevant ones, the ROC curve seems to overestimate the performance of several retrieval methods [6,7]. This is why an alternative method is used: the precision–recall (PR) curve [7,8].

At a fixed prevalence, Davis and Goadrich [7] have shown that there was a one-to-one correspondence between the ROC curve and the PR curve. However, other authors have shown that the area under the ROC curve (AUC) and the area under the PR curve (AUPRC) are not equivalent but show very large differences [9]. In the case of rare events, some authors have recommended the use of the AUPRC instead of the AUC [7]; however, up to now, no comparisons are readily available.

Funding: None.

* Corresponding author. Tel.: 33-478-865-764; fax: 33-472-115-141.

E-mail address: delphine.maucourt-boulch@chu-lyon.fr (D. Maucourt-Boulch).

What is new?

- In case of low-prevalence diseases, the area under the ROC curve (AUC) may overestimate the performance of a biomarker.
- At low prevalence, the area under the precision–recall curve (AUPRC) provides a better agreement with the positive predictive value of a biomarker.
- The AUPRC should be preferred over the AUC for the evaluating uncommon or rare disease biomarkers.
- The AUPRC should not be compared between populations with different disease prevalences because their values are prevalence dependent.

The present article compares the abilities of the AUC and the AUPRC in summarizing the classification performance of a biomarker according to the disease prevalence.

2. Methods**2.1. The ROC curve**

The ROC curve is a graphical technique used to assess the diagnostic accuracy of a continuous biomarker. It displays a trade-off between the sensitivity and the specificity of the biomarker over all possible biomarker threshold values. Using the conventional abbreviations (TP, FP, FN, and TN for the number of true positive, false positive, false negative, and true negative subjects), sensitivity equals $TP/(TP + FN)$ and specificity equals $FP/(TN + FP)$.

The AUC is a summary index of the biomarker performance. It ranges from 0.5 to 1 and corresponds to the probability that biomarker values from a randomly selected pair of diseased and nondiseased subjects are correctly ordered [3]. In this technique, sensitivity and specificity are unaffected by the disease prevalence and so is the AUC [6].

2.2. The precision–recall (PR) curve

The PR curve is an alternative approach for assessing the performance of a biomarker. It displays the trade-off between precision (instead of specificity) and sensitivity (also called recall) over all possible biomarker threshold values. Precision is the ratio $TP/(TP + FP)$, which corresponds to the PPV in the ROC approach. The PR curve focuses on the ability of the biomarker to identify diseased subjects; it ignores the correctly classified healthy subjects (TN), which is the dominant group in case of low-prevalence disease. Unlike a ROC curve, a PR curve is not necessarily monotonic across all biomarker thresholds because an increase in the threshold can decrease TP or FP.

The AUPRC is a summary statistic that reflects the ability of a biomarker to identify the diseased group. Denoting by x a biomarker value taken from the biomarker distribution in the diseased group, the AUPRC can be interpreted as the expectation, over all the possible x values, of the proportion of diseased subjects among all those whose biomarker value exceed x [8]. The values of the AUPRC range from 0 to 1, but whereas the expected value for random guessing of the AUC is 0.5, that of the AUPRC is prevalence dependent and tends to 0 when the prevalence decreases. Details on the estimation of the AUPRC and its confidence interval (with R codes) can be found in a recent article by Boyd et al. [8].

2.3. The simulation study

Simulations were performed on biomarker values that follow normal distributions in diseased and nondiseased subjects. The variance of the nondiseased group was fixed to $\sigma_1 = 1$, and the variance of the diseased group could range from $\sigma_2 = 0.01$ to 10. The difference between the means ($\mu_2 - \mu_1$) was allowed to range from 0 to 5 by 0.25 increments (21 steps). The size of the nondiseased group (n_1) was fixed at 10,000, but the size of the diseased group (n_2) could range from 100 (prevalence: 0.0099) to 10,000 (prevalence: 0.5).

For each n_2 , σ_2 , and ($\mu_2 - \mu_1$) combination, 1,000 sets of biomarker values in diseased and nondiseased subjects were generated. Then, at each prevalence value and at each variance of the diseased group, Pearson correlation coefficients (ρ) were computed between the 21×1000 AUC and AUPRC paired values obtained over all ($\mu_2 - \mu_1$) differences. ρ were also computed between AUC and PPV paired values as well as between AUPRC and PPV paired values.

2.4. Biomarker performance assessment

The biomarker performance was evaluated using the PPV and the NPV values calculated at the threshold at which the proportion of subjects classified as diseased is equal to the disease prevalence; that is, once the subjects are ranked in ascending biomarker values, the first n_1 subjects are considered as nondiseased and the remaining n_2 subjects as diseased. The PPV corresponds to the probability for a biomarker-positive subject to have the disease, whereas the NPV corresponds to the probability for a biomarker-negative subject to be actually nondiseased. Then, the AUC or the AUPRC should not be high for low values of the PPV or the NPV whatever the biomarker threshold.

3. Results

Table 1 lists that, at prevalence 0.5, the correlation between the AUC and the AUPRC is excellent ($\rho = 0.990$); besides, the AUC and the AUPRC reflected correctly the classification performance of the biomarker (correlations

Table 1. Pearson correlations coefficients between the AUC, the AUPRC, and the PPV at various prevalence values but a variance of the diseased group fixed at 0.5

Prevalence	ρ (AUC and AUPRC)	ρ (AUC and PPV)	ρ (AUPRC and PPV)
0.5	0.990	0.986	0.995
0.33	0.972	0.983	0.994
0.25	0.956	0.977	0.993
0.17	0.930	0.959	0.994
0.091	0.884	0.914	0.995
0.048	0.832	0.858	0.996
0.02	0.760	0.780	0.996
0.0099	0.706	0.724	0.995

Abbreviations: AUC, area under the ROC curve; AUPRC, area under the precision–recall curve; ROC, receiver operating characteristic; PPV, positive predictive value.

with PPV or NPV >0.95). The correlation between the AUC and the AUPRC decreased along with the prevalence, it dropped to 0.706 at prevalence 0.0099. The AUC and the AUPRC had close values at prevalence close to 0.5 but discrepant values as prevalence decreased. Similar results were obtained after varying the values of σ_2 (see [Supplementary Material](#) at www.jclinepi.com).

Fig. 1 shows that the AUPRC and the PPV decrease with the decrease of the prevalence and that the AUPRC and the NPV decrease with the decrease of $(\mu_2 - \mu_1)$. The decreases of the AUPRC and the PPV had the same profiles, which is confirmed by a correlation >0.95 . This correlation was almost constant over all simulations. In contrast, the AUC profile did not reflect the decrease of the PPV with the decrease of the prevalence; the correlation between the AUC and the PPV dropped from 0.986 at prevalence 0.5 to 0.724 at prevalence 0.0099 (**Table 1**).

Fig. 2 shows the differences between the AUC and the AUPRC values at two different prevalence values. With prevalence 0.5 (upper row), the biomarker is highly discriminant because of the limited overlap of its distributions in diseased and nondiseased subjects. This is confirmed by very high AUC and AUPRC (>0.95). With prevalence 0.0099 (lower row), a large number of nondiseased but a small number of diseased subjects had similar biomarker values. Here, the PPV was <0.25 at all thresholds, and the AUPRC was around 0.15 because of the high number of FPs vs. the TPs. However, the AUC remained very high [0.963 (95% confidence interval: 0.956, 0.971)]. **Fig. 2**, lower row, showed large discrepancies

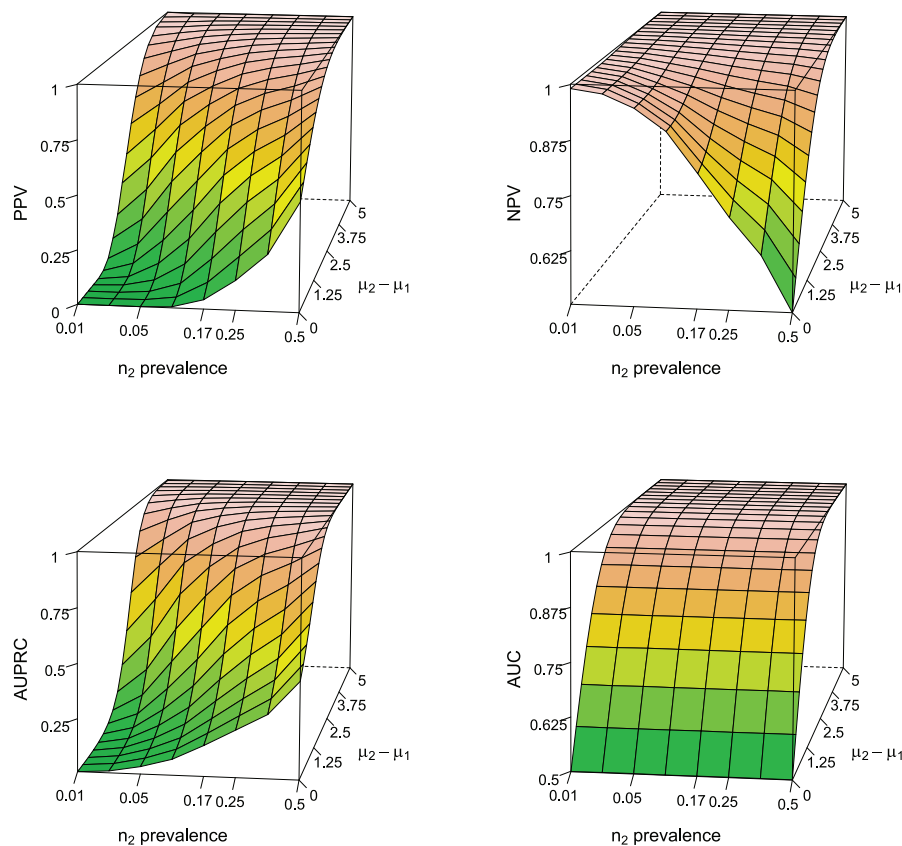


Fig. 1. Median values of the PPV, NPV, AUC, and AUPRC at variance of the biomarker values in the diseased group = 0.5. The X-axis indicates the prevalence of the disease in the diseased group. The Y-axis indicates the difference between the mean biomarker values in the diseased and non-diseased groups. PPV, positive predictive value; NPV, negative predictive value; AUC, area under the ROC curve; AUPRC, area under the precision–recall curve; ROC, receiver operating characteristic.

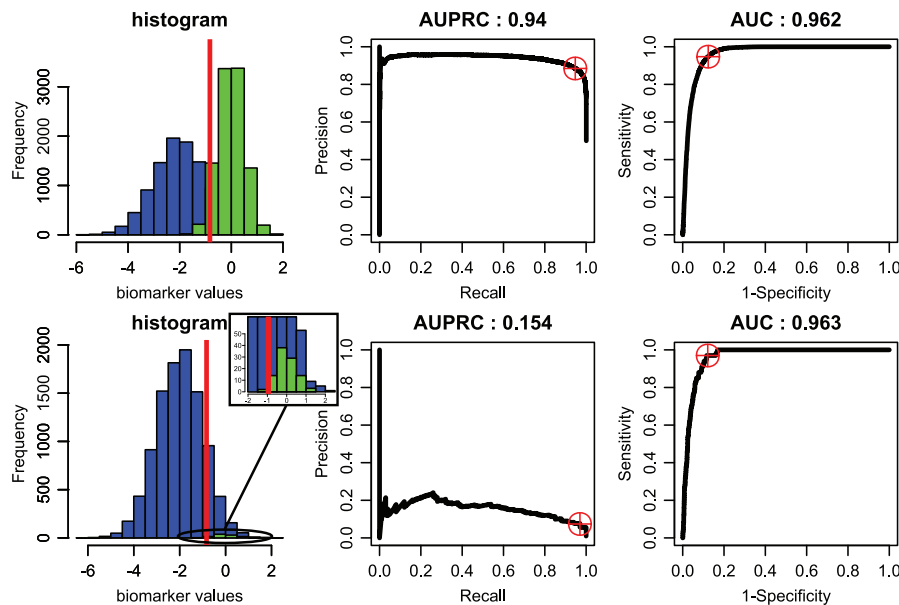


Fig. 2. Biomarker distribution, PR curve, and ROC curve with mean difference between biomarker values in the diseased and nondiseased groups = 2, variance of the biomarker values in the diseased group = 0.5, and disease prevalence = 0.5 in the upper row and 0.0099 in the lower row. The gray (red for the web version) vertical line on the histograms and the gray (red for the web version) circle on the AUPRC and AUC graphs indicate the threshold that minimizes the misclassification error considering disease prevalence = 0.5. Nota: The X-axis carries fictitious biomarker values. Precision = PPV and recall = sensitivity. PR, precision–recall; ROC, receiver operating characteristic; AUC, area under the ROC curve; AUPRC, area under the precision–recall curve; PPV, positive predictive value.

between the AUC (reflect of sensitivity and specificity) and the AUPRC (reflect of precision, or PPV, and recall—or sensitivity). For instance, at threshold -0.830 , recall (or sensitivity) and specificity were both high (respectively, 0.970 and 0.879) despite a low precision (or PPV = 0.074).

4. Discussion

The simulations indicated that, in low-prevalence diseases, the AUC may give an overoptimistic view of the performance of a diagnostic test; the AUPRC should be then preferred because, over different prevalence values, it gives coherent results with the PPV and the NPV.

However, the simulations did not show a clear-cut prevalence value below which the AUPRC should be preferred to the AUC. Thus, we may recommend to use both the AUC and the AUPRC but to prefer the AUPRC as the prevalence decreases, especially below 5%. Yet, the AUPRC is prevalence dependent and should not be compared between populations with different disease prevalence values. In the latter case, it is the AUC that should be used. For the same reason, the AUPRC has a limited interest when the prevalence in the sample under study does not reflect the prevalence in the population of interest (typically, in the early phases of clinical trials). Besides, it should be kept in mind that a higher AUC does not necessarily imply a higher AUPRC and conversely [7].

A reliable index of diagnostic test performance in low-prevalence diseases seems to be still lacking. In a

comparative study of predictive models for adverse drug reaction in a context of low-prevalence (around 0.006) [10], the authors reported improvements in the detection of adverse drug reactions that were severely underestimated by the AUC but detected by precision and recall. Thus, the use of the AUPRC together with the AUC should ensure more accurate model evaluations and comparisons. This has been recently done in the field of medical imaging where the prevalence of diseased voxels after stroke is usually low. The AUC gave a biased view of the model predictive performance, whereas the AUPRC reflected better the quality of the prediction of the diseased volumes [9].

In our simulations, the parameters (prevalence, mean, and variance) were varied independently although this may not be true in practice [11]. However, this allowed considering all possible scenarios. One limit of the present checking process is that the PPV and the NPV are threshold dependent. We have chosen to adapt the biomarker threshold to the prevalence although other strategies for defining the optimal threshold could have been considered, for example, maximizing a utility function [12]. Nevertheless, in some simulations, high AUC values were obtained with low PPV values at all thresholds; this is one crucial limit of the AUC in low-prevalence diseases.

Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jclinepi.2015.02.010>.

References

- [1] The TDR Diagnostics Evaluation Expert Panel, Banoo S, Bell D, Bossuyt P, Herring A, Mabey D, Poole F, et al. Evaluation of diagnostic tests for infectious diseases: general principles. *Nat Rev Microbiol* 2010;1:17–29.
- [2] Grimes DA, Schulz KF. Uses and abuses of screening tests. *Lancet* 2002;359:881–4.
- [3] Pepe MS. The statistical evaluation of medical tests for classification and prediction. Oxford: Oxford University Press; 2003.
- [4] Quemeneur AS, Trocetto JM, Ea HK, Woimant F, Lioté F. Miscellaneous non-inflammatory musculoskeletal conditions. Musculoskeletal conditions associated with Wilson's disease. *Best Pract Res Clin Rheumatol* 2011;25(5):627–36.
- [5] Brott T, Marler JR, Olinger CP, Adams HP, Tomsick T Jr, Barsan WG, et al. Measurements of acute cerebral infarction: lesion size by computed tomography. *Stroke* 1989;20:871–5.
- [6] Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* 2006;27:861–74.
- [7] Davis J, Goadrich M. The relationship between precision-recall and ROC curves. In: *Proceedings of the 23rd International Conference on Machine Learning*. New York: ACM; 2006:233–40.
- [8] Boyd K, Eng K, Page C. Area under the precision-recall curve: Point estimates and confidence intervals. In: Blockeel H, Kersting K, Nijssen S, Železný F, editors. *Machine learning and knowledge discovery in databases*. Springer: Berlin Heidelberg; 2013:451–66. Volume 8190 of *Lecture Notes in Computer Science*.
- [9] Bouts MJ, Tiebosch IA, Van der Toorn A, Viergever MA, Wu O, Dijkhuizen RM. Early identification of potentially salvageable tissue with MRI-based predictive algorithms after experimental ischemic stroke. *J Cereb Blood Flow Metab* 2013;33:1075–82.
- [10] Liu M, Wu Y, Chen Y, Sun J, Zhao Z, Chen X, et al. Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *J Am Med Inform Assoc* 2012;19:e28–35.
- [11] Knottnerus JA, Leffers P. The influence of referral patterns on the characteristics of diagnostic tests. *J Clin Epidemiol* 1992;45:1143–54.
- [12] Pauker SG, Kassirer JP. Therapeutic decision making: a cost-benefit analysis. *N Engl J Med* 1975;293:229–34.

Comparison between the AUC and the AUPRC at various variances of the biomarker values in the diseased group.

Table A – Pearson correlation between the PPV and each of the AUC and the AUPRC at various values of the prevalence P of the disease.

σ_1	σ_2	P = 0.5			P = 0.25			P = 0.091			P = 0.0099		
		$\rho(\text{AUC/PPV})$	$\rho(\text{AUPRC/PPV})$	$\rho(\text{AUC/PPV})$	$\rho(\text{AUC/PPV})$	$\rho(\text{AUPRC/PPV})$	$\rho(\text{AUC/PPV})$	$\rho(\text{AUC/PPV})$	$\rho(\text{AUPRC/PPV})$	$\rho(\text{AUC/PPV})$	$\rho(\text{AUC/PPV})$	$\rho(\text{AUPRC/PPV})$	$\rho(\text{AUPRC/PPV})$
1	.01	1.000	0.967	0.981	0.96	0.96	0.838	0.979	0.609	0.985			
1	.05	1.000	0.974	0.983	0.965	0.965	0.846	0.984	0.615	0.988			
1	0.1	0.998	0.981	0.985	0.97	0.97	0.855	0.987	0.625	0.992			
1	0.5	0.986	0.995	0.977	0.993	0.993	0.914	0.995	0.724	0.995			
1	1	0.984	0.984	0.976	0.993	0.993	0.949	0.996	0.849	0.996			
1	2	0.994	0.992	0.991	0.996	0.996	0.983	0.998	0.946	0.997			
1	5	1.000	0.999	0.999	0.999	0.999	0.995	0.999	0.948	0.988			
1	10	1.000	0.999	0.998	0.999	0.999	0.991	0.996	0.887	0.95			

σ_1 : variance of the biomarker values in the non-diseased group - σ_2 : variance of the biomarker values in the diseased group.

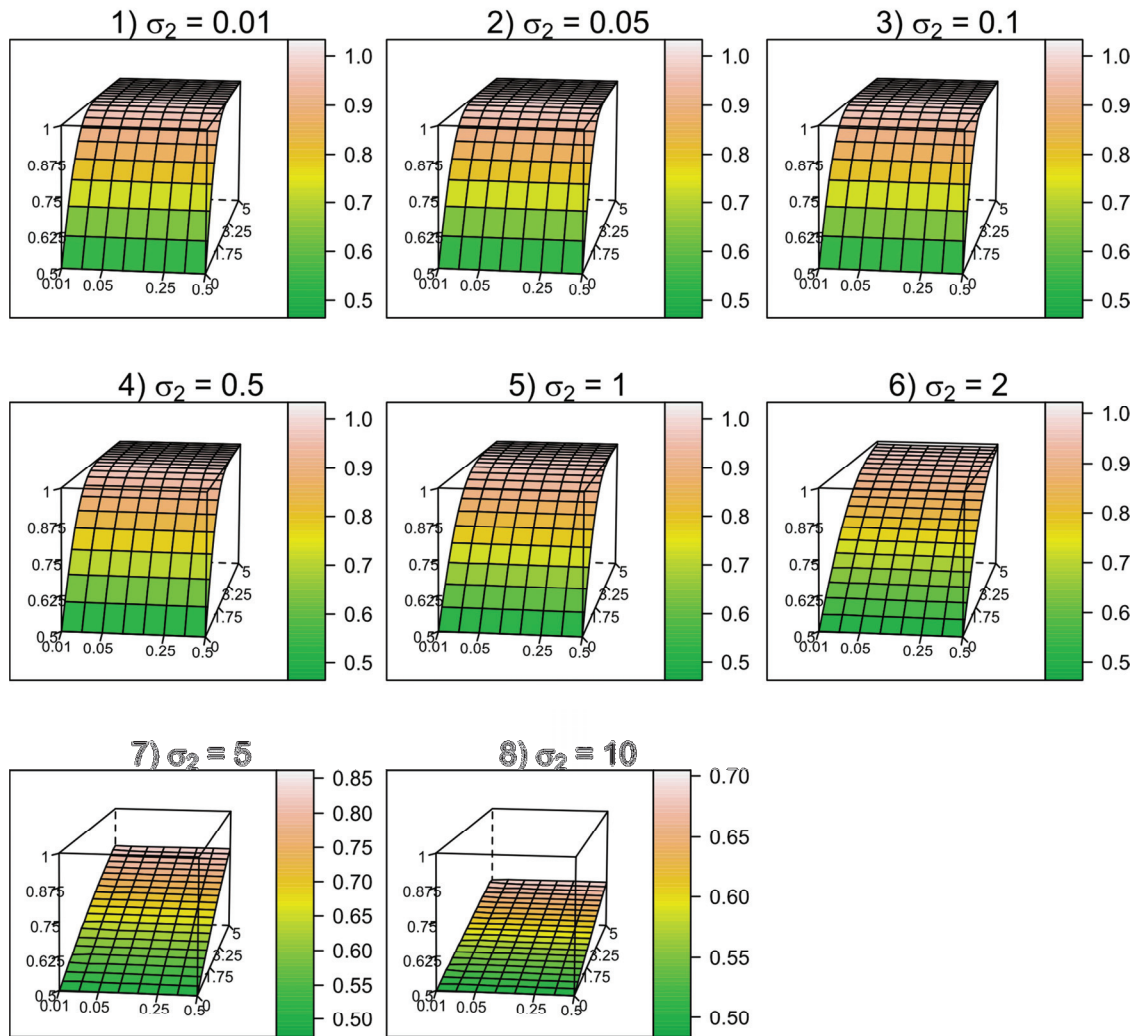


Figure A - Median AUC values over the 1000 sets of biomarker values in diseased and non-diseased subjects. Each subfigure corresponds to a simulation made at a specific value of σ_2 , the variance of the biomarker values in the diseased group. The X axis indicates the prevalence of the disease in the diseased group. The Y axis indicates the difference between the means of the biomarker values in the diseased and the non-diseased group. The Z axis indicates the value of the AUC.

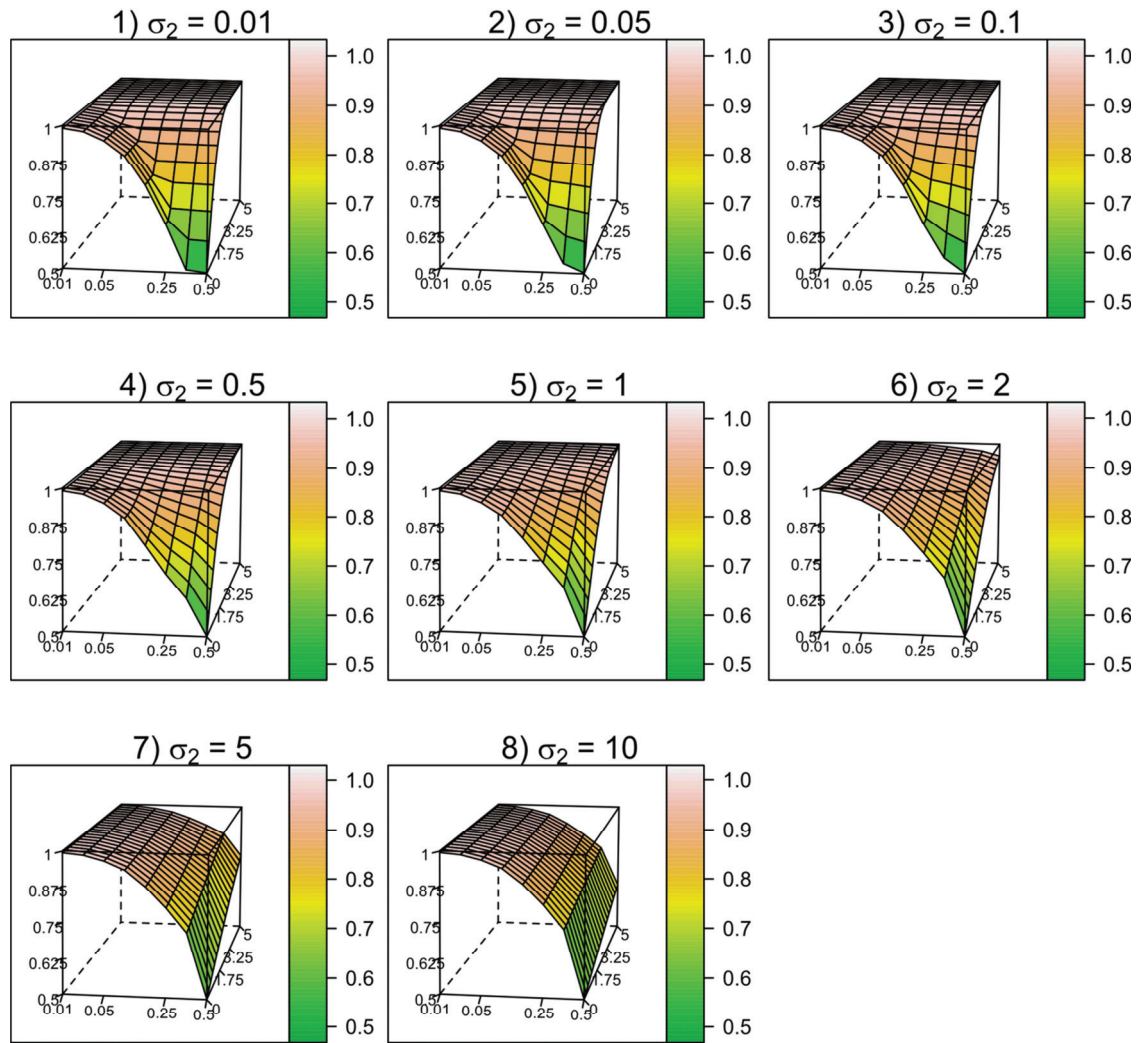


Figure B - Median NPV values over the 1000 sets of biomarker values in diseased and non-diseased subjects. Each subfigure corresponds to a simulation made at a specific value of σ_2 , the variance of the biomarker values in the diseased group. The X axis indicates the prevalence of the disease in the diseased group. The Y axis indicates the difference between the means of the biomarker values in the diseased and the non-diseased group. The Z axis indicates the value of the NPV.

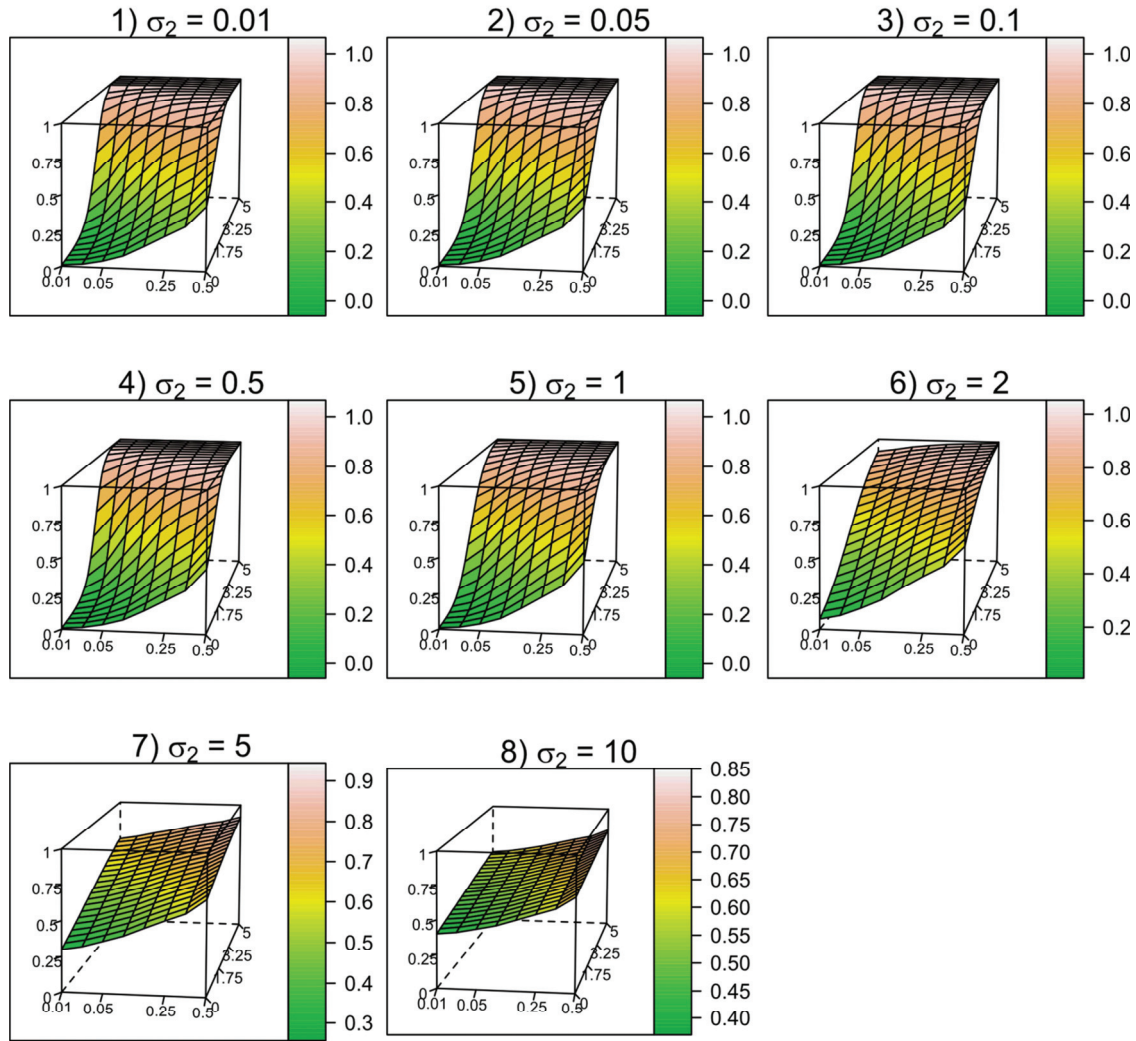


Figure C - Median AUPRC values over the 1000 sets of biomarker values in diseased and non-diseased subjects. Each subfigure corresponds to a simulation made at a specific value of σ_2 , the variance of the biomarker values in the diseased group. The X axis indicates the prevalence of the disease in the diseased group. The Y axis indicates the difference between the means of the biomarker values in the diseased and the non-diseased group. The Z axis indicates the value of the AUPRC.

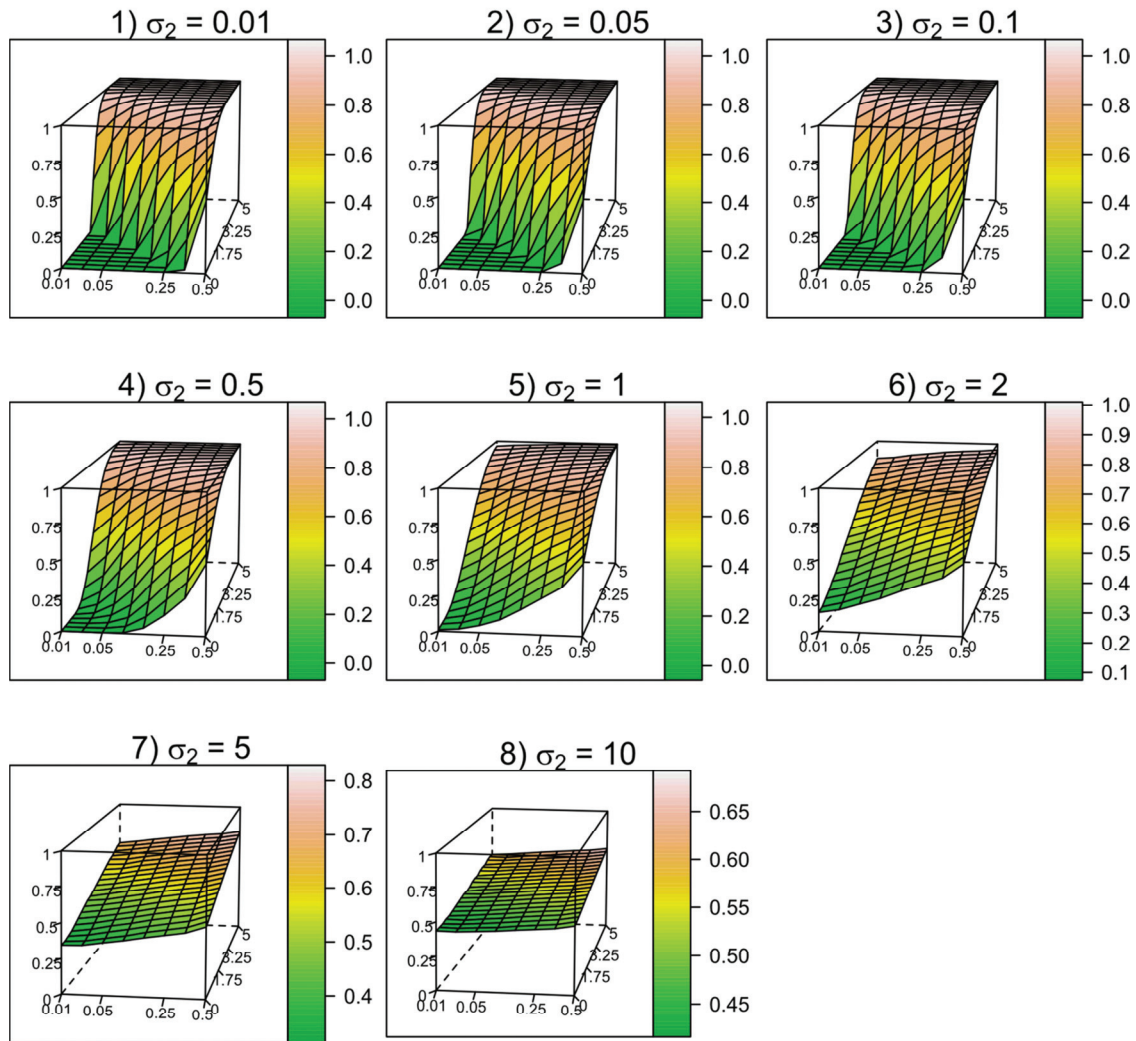


Figure D - Median PPV values over the 1000 sets of biomarker values in diseased and non-diseased subjects. Each subfigure corresponds to a simulation made at a specific value of σ_2 , the variance of the biomarker values in the diseased group. The X axis indicates the prevalence of the disease in the diseased group. The Y axis indicates the difference between the means of the biomarker values in the diseased and the non-diseased group. The Z axis indicates the value of the PPV.

Extensions

L'article est centré sur la comparaison entre l'AUPRC et l'AUC qui sont des mesures de performance globales. Néanmoins si l'on est intéressé à évaluer les performances à un seuil donné, on peut montrer qu'il faut préférer l'indice de Jaccard au pourcentage de bien classés lorsque la prévalence est faible. Les raisons sont analogues à celles exprimées dans l'article : le pourcentage de bien classés intègre le nombre de vrais négatifs tandis que l'indice de Jaccard l'ignore. Or pour de faibles prévalences, les vrais négatifs sont prépondérants et vont masquer les variations de faux négatifs et de faux positifs. De même que l'AUC, le pourcentage d'observations bien classées par le modèle peut alors être excellent (par exemple $AUC=0,79$ et $\%BC=0,83$ - NGUYEN et al., 2008) alors qu'aux seuils correspondant à une spécificité élevée (exemple : $Sp=0,975$) la sensibilité reste modérée (de l'ordre de 0,5). Pour un AVC de 30 mL (sur 675 mL de tissu) cela ferait 15 mL de vrais positifs, 16,125 mL de faux positifs et 15 mL de faux négatifs : le modèle est donc peu fiable.

Il est aussi possible d'évaluer l'intérêt de l'AUC partielle. Pour cela, nous avons choisi des gammes de spécificité du type $(Sp_{min}, 1)$. En augmentant convenablement le seuil de spécificité minimale avec la prévalence, l'AUC partielle peut donner une évaluation proche de celle obtenue avec l'AUPRC (voir figure II.3). Mais comme mentionné précédemment, le choix du seuil reste très subjectif ce qui n'est pas souhaitable.

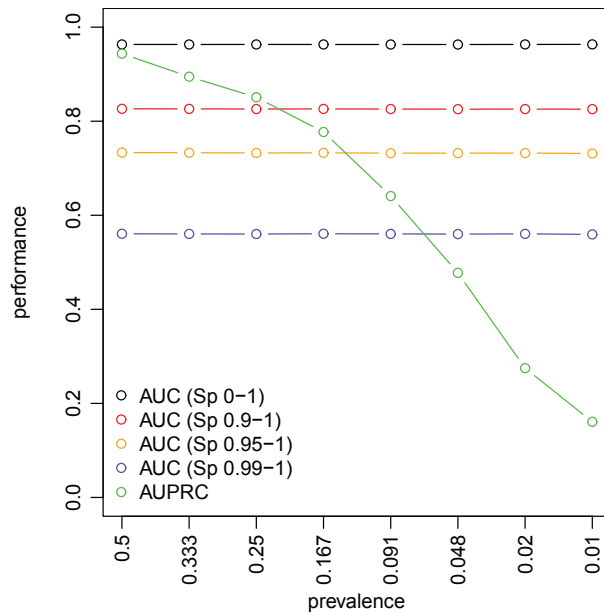


FIGURE II.3 – Évolution de l'AUC (en noir), des AUC partielles pour $Sp_{min} = 0,9$ (en rouge), $Sp_{min} = 0,95$ (en orange), $Sp_{min} = 0,99$ (en bleu) et de l'AUPRC (en vert) avec la prévalence en reprenant les mêmes paramètres de distribution du biomarqueur que pour la figure 2 de l'article page 32.

2.5 Critères retenus pour l'évaluation des modèles prédictifs

Le principal critère d'évaluation du modèle prédictif au niveau voxel sera donc **l'AUPRC**. Néanmoins l'AUC sera aussi reportée pour pouvoir se situer vis-à-vis des autres études.

L'évaluation volumique se fera en considérant **l'erreur volumique en valeur absolue** ($|\hat{V} - V|$). Elle est préférée à la corrélation et à l'indice de Jaccard parce qu'elle est facilement interprétable. Une erreur volumique relative à la taille de la lésion pourrait aussi être considérée mais elle risquerait de pénaliser trop fortement les prédictions pour les petits AVC. Les critères volumiques pour l'AVC sont généralement de l'ordre d'une dizaine de mL : 10 mL pour la reperfusion ou le mismatch dans DEFUSE (ALBERS et al., 2006), 15 mL pour le mismatch dans DEFUSE 2 (LANSBERG et al., 2012). Une précision volumique des prédictions de l'ordre du mL est donc nécessaire.

3 Comparaison de modèles prédictifs pour l'AVC

Depuis le début des années 2000, un intérêt croissant est porté à la modélisation prédictive pour l'AVC. En effet, par rapport à la séquence en diffusion où un simple seuillage permettait d'identifier la zone nécrosée, la séquence en perfusion s'est avérée nettement plus difficile à exploiter. De nombreux travaux ont cherché à identifier le seuil et le paramètre de perfusion optimaux pour définir la zone de pénombre (GRANDIN et al., 2002; MOURIDSEN et al., 2004; CHRISTENSEN et al., 2009; OLIVOT et al., 2009a; CARRERA et al., 2011; FORKERT et al., 2013, etc). Même si ces travaux ont montré la supériorité de certains paramètres de perfusion et de certaines méthodes de traitement du signal, aucun n'a mis en évidence un seuil de paramètre permettant une identification fiable de la zone de pénombre. Ce résultat n'est pas surprenant au vu des conséquences complexes de l'AVC, et de leur caractérisation indirecte et entachée d'incertitudes par l'IRM. Considérer qu'un seuil de paramètre IRM va suffir à distinguer les voxels sains des voxels à risque de nécrose semble aujourd'hui peu réaliste. Certes l'amélioration des méthodes de mesure et de traitement du signal (ZAHARCHUK et al., 2009) ainsi que l'utilisation de nouveaux paramètres IRM (MOURIDSEN et al., 2014) se révéleront probablement bénéfiques. Mais il apparaît aujourd'hui nécessaire de considérer une approche prédictive multiparamétrique afin d'exploiter au mieux l'ensemble de l'information issue de l'examen IRM (ØSTERGAARD et al., 2009; WU et al., 2010).

3.1 Les approches prédictives pour l'AVC

État de l'art

De nombreuses approches prédictives ont été proposées au cours des dix dernières années. Nous ne nous intéresserons ici qu'aux approches voxel à voxel c'est-à-dire prédisant le devenir des tissus pour chaque voxel. Les approches prédisant directement le volume de lésion final à partir de volumes de perfusion, de diffusion et des données cliniques ne sont donc pas considérées ici. En agrégeant les données, ces approches volumiques ne permettent qu'une modélisation grossière et sont par construction limitées : il est impossible d'exploiter l'information spatiale ou même de tenir compte des interactions entre paramètres IRM.

Les principaux travaux prédictifs sont reportés dans le tableau II.4 et peuvent être divisés en 3 catégories. Les **méthodes de reconnaissance de groupes non supervisée** qui analysent l'hétérogénéité des valeurs de paramètres IRM pour former des groupes de voxels (sains, à risque, nécrosés). Cette approche, développée par exemple dans les articles de CARANO et al., 2000, et de SHEN et al., 2008a, utilise uniquement les données d'IRM mesurées à l'admission du patient. Elle n'est donc pas affectée par les erreurs ou approximations faites lors de la délimitation de la lésion finale. Des **méthodes d'analyse d'images** ont aussi été proposées. L'article de ROSSO et al., 2009, propose une méthode de croissance de région qui mime l'expansion de l'AVC en minimisant progressivement un terme d'énergie. Celui-ci régule l'homogénéité de l'intensité dans la zone de nécrose ainsi que la forme de la lésion. Toutefois, cette approche a

été développée pour des contextes où l'on ne dispose que des données en diffusion. Or, même si le terme d'homogénéité peut être adapté au cas multiparamétrique, sa définition est délicate puisqu'elle implique de fixer un certain nombre de paramètres. Une méthode alternative a été proposée dans l'article de SCALZO et al., 2012, pour exploiter l'aspect régional de l'AVC. Il s'agit non plus de considérer chaque voxel mais plutôt sa région sous forme d'un cube. Chaque cube est réorienté pour être comparable aux autres, puis un classificateur est entraîné sur un échantillon du jeu de données à repérer les cubes qui vont évoluer vers la nécrose. Pour cela, une régression spectrale à noyau (CAI et al., 2007) est utilisée. La régression spectrale permet d'imposer des prédictions similaires à des voxels proches et l'utilisation d'un noyau permet d'améliorer la capacité discriminante de la régression. Cette approche contient plusieurs idées intéressantes : prise en compte de l'information régionale, devenir similaire entre voxels proches et non-linéarité des risques. Cependant elle n'a pour le moment été proposée que dans un cadre univarié.

Enfin, la majorité des travaux utilisent des **méthodes d'apprentissage supervisé** avec principalement le modèle de régression logistique (WU et al., 2006) et ses extensions (WU et al., 2004 ; NGUYEN et al., 2008 ; KIDWELL et al., 2013) ainsi que des méthodes d'apprentissage automatique. Ces méthodes plus complexes et plus coûteuses numériquement permettent une utilisation plus souple des variables prédictives, supposées indépendantes et d'effet linéaire en régression. Ainsi les réseaux de neurones (BAGHER-EBADIAN et al., 2011 ; HUANG et al., 2010) et les machines à vecteurs de support (HUANG et al., 2011) ont montré des performances intéressantes. Cependant, la seule étude comparative disponible pour le moment n'a pas montré de supériorité de ces méthodes vis-à-vis de la régression logistique (BOUTS et al., 2013). Cette étude a été réalisée sur des données animales et l'extrapolation de ces résultats chez l'homme n'a rien d'évident (CASALS et al., 2011). Enfin, une seule étude s'est intéressée à la prise en compte de la variabilité inter-patients dans l'approche prédictive (KIDWELL et al., 2013). La méthodologie statistique utilisée n'est toutefois pas complètement claire : un modèle mixte a été ajusté mais les auteurs ne précisent pas comment les prédictions étaient calculées à partir de ce modèle (probablement à l'aide du modèle marginal).

Objectifs

Nous nous proposons de clarifier l'apport des différents modèles prédictifs proposés dans la littérature en comparant leurs performances prédictives. Le modèle logistique sera pris comme modèle de référence.

Article	Données	Paramètres IRM (pré-traitement)	Méthode(s) évaluée(s)	Résultat(s)
Méthodes de reconnaissance de groupes non supervisée				
CARANO et al., 2000	21 rats	DWI, FA, T2, PD, CBF	Méthode des K moyennes (KM) Méthode des C moyennes floues (FCM)	KM>FCM %BC=95.9% $\rho(V, \hat{V}) = 0,97$
LU et al., 2005	45 patients	ADC, T2 (lissage non-linéaire)	ISODATA	$\rho(V, \hat{V}) = 0,70$
SHEN et al., 2008a	36 rats	ADC, CBF	ISODATA, fréquence spatiale de nécrose	AUC \approx 0,9
Méthodes d'analyse d'images				
ROSSO et al., 2009	98 patients	ADC	Modèle de croissance de région	$\rho(V, \hat{V}) = 0,83$
SCALZO et al., 2012	25 patients	ADC or TMAX	Régression spectrale (SR-KDA) Régression linéaire (LIN)	SR-KDA>LIN AUC \approx 0,909
Méthodes d'apprentissage supervisé				
BAGHER-EBADIAN et al., 2011	12 patients	DWI, T2, PD, T1 (lissage, normalisation)	ANN (sigmoïde, 2 couches, 6 neurones)	AUC de 0,89
BOUTS et al., 2013	19 rats	T2, ADC, CBF, MTT, TMAX (normalisation)	LR, GAM (cubic spline regression), SVM (Laplace, C=1,2, $\sigma=0,39$), ADA (profondeur=2, arbres=125), RF (try=1, arbres=200)	non-infériorité de LR AUC $\in [0,8; 1]$ AUPRC $\in [0,4; 0,8]$

GOTTRUP et al., 2005	14 patients	ADC, DWI, T2, CBF, CBV, MTT (normalisation)	K plus proches voisins (avec différentes distances)	AUC = 0,814±0,001
HUANG et al., 2011	36 rats	ADC, CBF + information spatiale	ANN (sigmoïde, 1 couche, 6 neurones) SVM (Gaussien, C et σ optimisés)	SVM>ANN AUC=0,88±0,028
KIDWELL et al., 2013	66 patients	ADC, TMAX, MTT, CBF, CBV (normalisation)	MLR	%BC = 0,71 (AUC=0,76) critère de reperfusion
NGUYEN et al., 2008	74 patients	T2, ADC, DWI, CBF, CBV, MTT (normalisation)	LR, SAR	SAR>LR AUC=0,79
WU et al., 2001	94 patients	T2, ADC, DWI, CBF, CBV, MTT (normalisation)	LR, seuillage	LR \approx seuillage
WU et al., 2004	12 patients	T2, ADC, DWI, CBF, CBV, MTT, TMAX (normalisation)	LR stratifié sur le type de tissu LR	LR stratifié > LR
WU et al., 2006	38 patients	T2, ADC, DWI, CBF, CBV, MTT, TMAX	LR	AUC = 0,9 $ \hat{V} - V = 37 \pm 35$ mL rt-PA \downarrow performances

TABLE II.4 – Synthèse des approches prédictives voxel à voxel proposées dans la littérature. Les indices utilisés pour évaluer ces approches sont reportés dans le tableau II.2.
FA : anisotropie du gradient en diffusion ; PD : Densité protonique ; SAR : modèle spatial autoregressif ; ISODATA : méthode des K moyennes modifiée.

Étant donné la multitude d'approches proposées, une étude comparative complète n'était pas réalisable. Nous nous sommes donc concentrés sur les méthodes d'apprentissage supervisé puisqu'elles montraient les performances les plus prometteuses. Le filtrage des données ainsi que la prise en compte de l'information spatiale et morphologique ayant été suggérés bénéfiques, nous évaluerons aussi leur incidence sur les performances du modèle. L'étude se déroulera en trois temps :

- **Évaluation de l'intérêt des méthodes d'apprentissage automatique** : ces méthodes étant très coûteuses en temps de calcul, une étude préliminaire permettra de ne retenir que la plus prometteuse.
- **Améliorations du modèle logistique standard** : en considérant le modèle logistique de WU et al., 2006, comme un modèle de référence, nous évaluerons l'impact des choix de modélisation, de variables prédictives et de filtrage des données.
- **Intérêt clinique de la modélisation prédictive** : à partir de l'approche prédictive retenue, nous montrerons comment les modèles prédictifs peuvent être utilisés pour (i) évaluer l'effet d'un traitement et (ii) former des critères d'éligibilité au traitement.

Avant de détailler l'étude effectuée, nous présentons dans la section suivante les traitements qui ont été utilisés pour rendre exploitable les images issues de l'IRM.

3.2 Traitement des images issues d'IRM

Le traitement des données d'imagerie a pour but d'exclure les voxels non pertinents du domaine d'étude, d'identifier les structures cérébrales sur les images issues de l'IRM, de normaliser les données entre les patients et éventuellement d'améliorer le rapport signal sur bruit des mesures effectuées par IRM. Le traitement se divise en quatre étapes qui sont présentées dans cette section et illustrées sur un patient. Afin de permettre la reproductibilité de ces étapes, une librairie de fonction a été développée et mise à disposition sur le CRAN : <http://cran.r-project.org/web/packages/MRIaggr/index.html>. Son fonctionnement sera décrit dans la seconde partie de la thèse, section 2 page 170.

Élimination du fond

L'IRM étant acquise avec un champ large elle contient non seulement les voxels correspondant au cerveau mais aussi ceux environnants : yeux, boîte crânienne, etc. Ces voxels sont éliminés en identifiant la zone correspondant au cerveau (figure II.4).

Identification des structures cérébrales

L'identification des hémisphères est effectuée en recherchant le plan maximisant la symétrie entre les deux hémisphères pour l'intensité T1 (voir figure II.5, image de gauche).

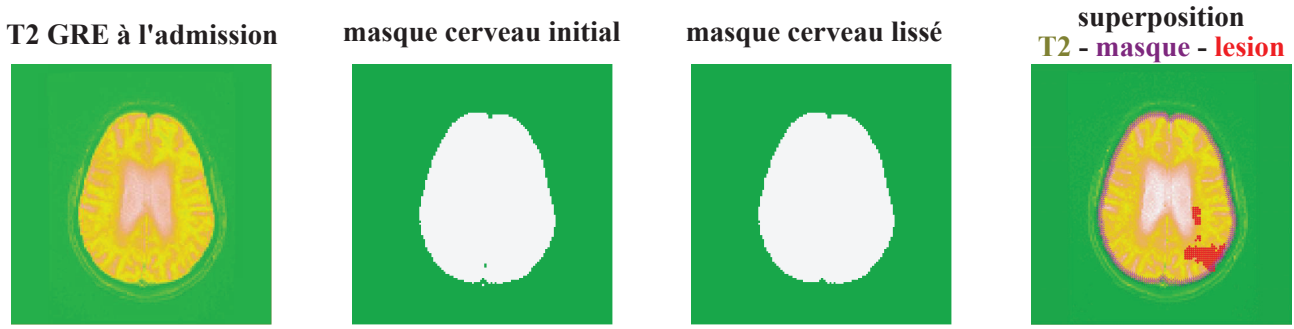


FIGURE II.4 – Identification des voxels correspondant au cerveau en deux étapes : algorithme des K moyennes puis retraitement du masque obtenu avec des critères de cohérence spatiale.

L'identification du contenu des voxels en substance blanche, substance grise et liquide cérébro-spinal est réalisée à l'aide d'un algorithme de segmentation probabiliste (FENG et al., 2012). Ce dernier fournit pour chaque voxel trois probabilités d'appartenance : chacune indique la probabilité que le voxel corresponde à de la substance grise, de la substance blanche ou du liquide cérébro-spinal (figure II.5, image du milieu).

Enfin la proximité spatiale à la lésion est calculée selon la distance euclidienne du voxel à la lésion :

$$distance(i) = \begin{cases} 0 & \text{si } i \in MASK_{H0} \\ d_E(i, MASK_{H0}) & \text{sinon} \end{cases}$$

où $MASK_{H0}$ est l'ensemble des voxels de la zone de lésion identifiée à l'admission du patient et d_E est la distance euclidienne. Elle est représentée figure II.5, image de droite.

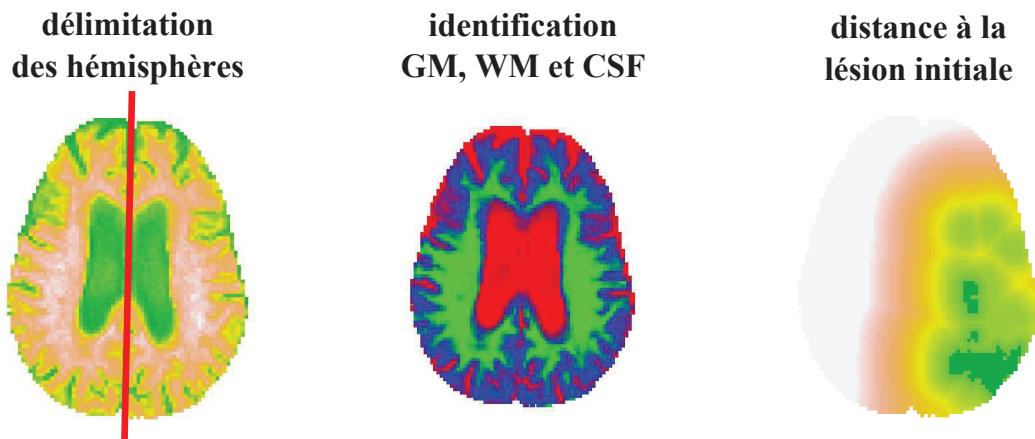


FIGURE II.5 – Identification de différentes structures cérébrales. De gauche à droite : plan sagittal moyen séparant les hémisphères (en rouge) ; substance blanche (en vert), substance grise (en bleu) et liquide cérébro-spinal (en rouge) ; distance à la lésion initiale avec en vert foncé la lésion initiale soit une distance nulle.

Filtrage des paramètres d'imagerie

L'application d'un filtre sur les paramètres IRM peut permettre d'améliorer le ratio signal sur bruit. Une première solution consiste à moyenner localement les valeurs : le signal étant supposé localement constant et le bruit aléatoire, cela va annuler le bruit et conserver le signal (**lissage gaussien**, première ligne de la figure II.6). Un des principaux défauts de cette approche est qu'elle va moyenner des signaux qui sont réellement différents. Pour cela une version robuste est aussi proposée : elle intègre uniquement les voxels d'intensité proche dans la moyenne (**lissage bilatéral** - TOMASI et al., 1998).

Un autre problème souvent rencontré avec les données de perfusion ainsi qu'avec le T2 est la présence d'artéfacts, c'est-à-dire de zones ayant des valeurs élevées de paramètre mais ne correspondant pas à l'AVC. Par exemple, sur la cartographie de TMAX représentée sur la figure II.6, le liquide cérébro-spinal ainsi que les tissus environnants ont un signal fort. La correction proposée consiste à soustraire voxel à voxel la valeur controlatérale (**soustraction du controlatéral**, seconde ligne de la figure II.6). Cette correction est cependant imparfaite à cause de la symétrie inexacte des structures cérébrales. Le résultat de ces trois filtres est présenté à la dernière ligne de la figure II.6.

Normalisation des données

La dernière étape de traitement des données est la normalisation. Les données étant multicentriques, les différences de réglages entre les IRM vont s'ajouter aux spécificités de chaque patient et rendre les valeurs peu comparables entre patients. Cela est particulièrement visible pour la séquence en diffusion (voir la première ligne de la figure II.7).

La normalisation des paramètres IRM pour chaque patient est effectuée vis-à-vis de l'hémisphère sain, liquide cérébro-spinal exclu (**normalisation intra-patient**). Pour chaque coupe la valeur moyenne de l'hémisphère controlatéral est calculée sur trois coupes glissantes (i.e. la coupe courante, celle du dessous et celle du dessus). L'écart-type est quant à lui calculé sur l'ensemble de l'hémisphère controlatéral. Les valeurs prises par chaque paramètre IRM sont alors centrées et réduites, avec la moyenne et l'écart-type précédemment calculés. Cette normalisation concerne les paramètres IRM bruts et lissés mais ne concerne pas ceux avec soustraction du controlatéral. Le choix du volume controlatéral (une coupe, trois coupes glissantes ou l'ensemble des coupes) dans le calcul de la moyenne et de l'écart-type a été effectué en optimisant la capacité du DWI à l'admission à identifier la lésion initiale. Utiliser trois coupes glissantes était un bon compromis entre une estimation trop locale qui donnait des intensités trop variables d'une coupe à l'autre, et l'utilisation du volume entier qui ne permettait pas de normaliser correctement les coupes basses.

Enfin pour pouvoir comparer les effets des variables prédictives et éviter des problèmes d'instabilité numérique dans les algorithmes d'estimation, tous les paramètres IRM étaient

centrés réduits. Le paramètre de distance à la lésion initiale était uniquement réduit. Les probabilités d'appartenance au liquide cérébro-spinal, à la substance blanche et à la substance grise n'ont pas été modifiées. Le tableau II.5 récapitule l'ensemble des étapes de normalisation.

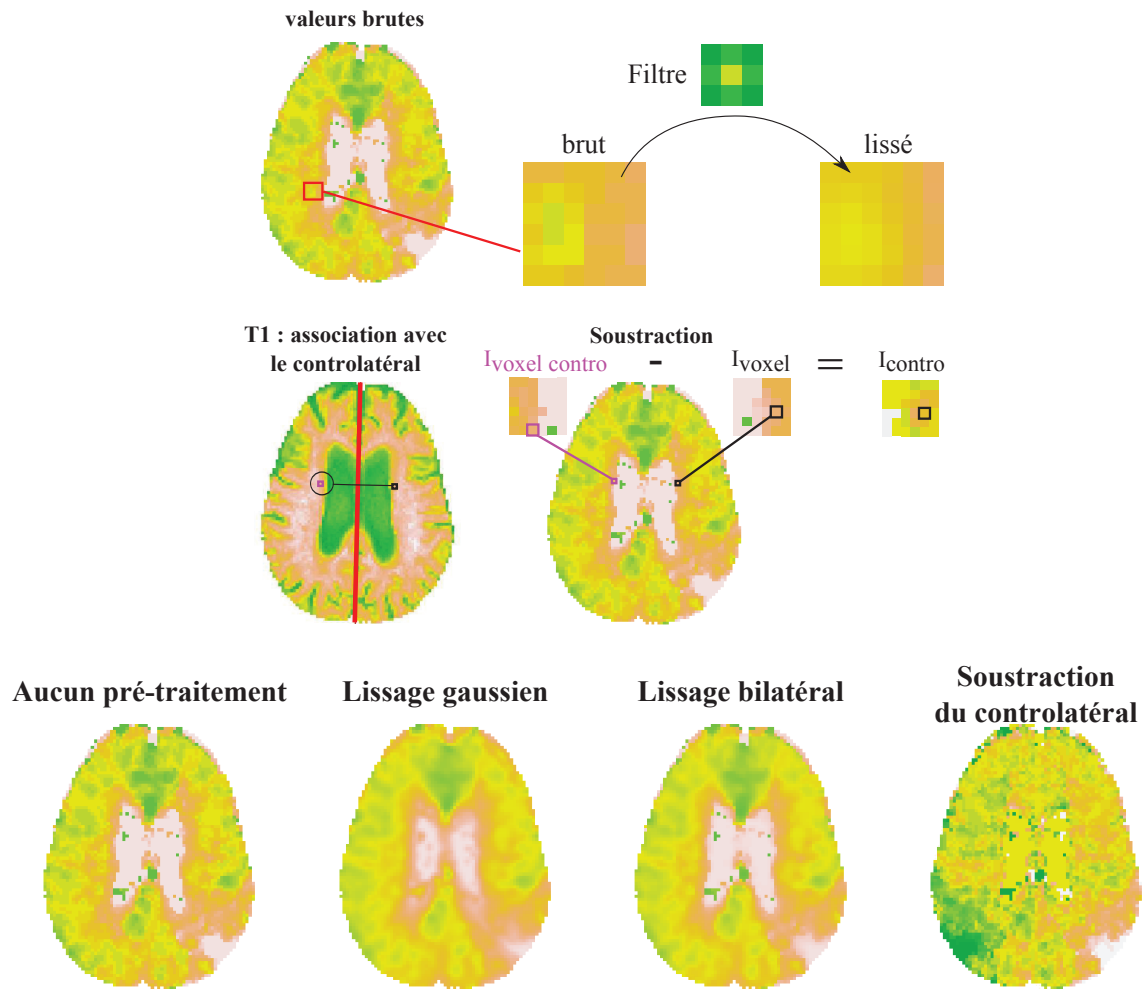


FIGURE II.6 – Filtrage du paramètre IRM nommé TMAX. En haut, principe du filtrage gaussien. Au milieu, principe de la soustraction du controlatéral. En bas résultats des différents filtrages.

paramètre	type de filtrage	normalisation intra-patient	centrage	réduction
X	Aucun	Oui	Oui	Oui
X	Lissage gaussien	Oui	Oui	Oui
X	Lissage gaussien bilatéral	Oui	Oui	Oui
X	Soustraction voxel à voxel de la valeur controlatérale	Non	Oui	Oui
distance	Aucun	Non	Non	Oui
GM, WM, CSF	Aucun	Non	Non	Non

TABLE II.5 – Étapes de normalisation associées à chaque paramètre. X renvoie à n'importe quel paramètre IRM parmi : DWI, ADC, T2 FLAIR, CBF, CBV, TMAX, MTT, TTP.

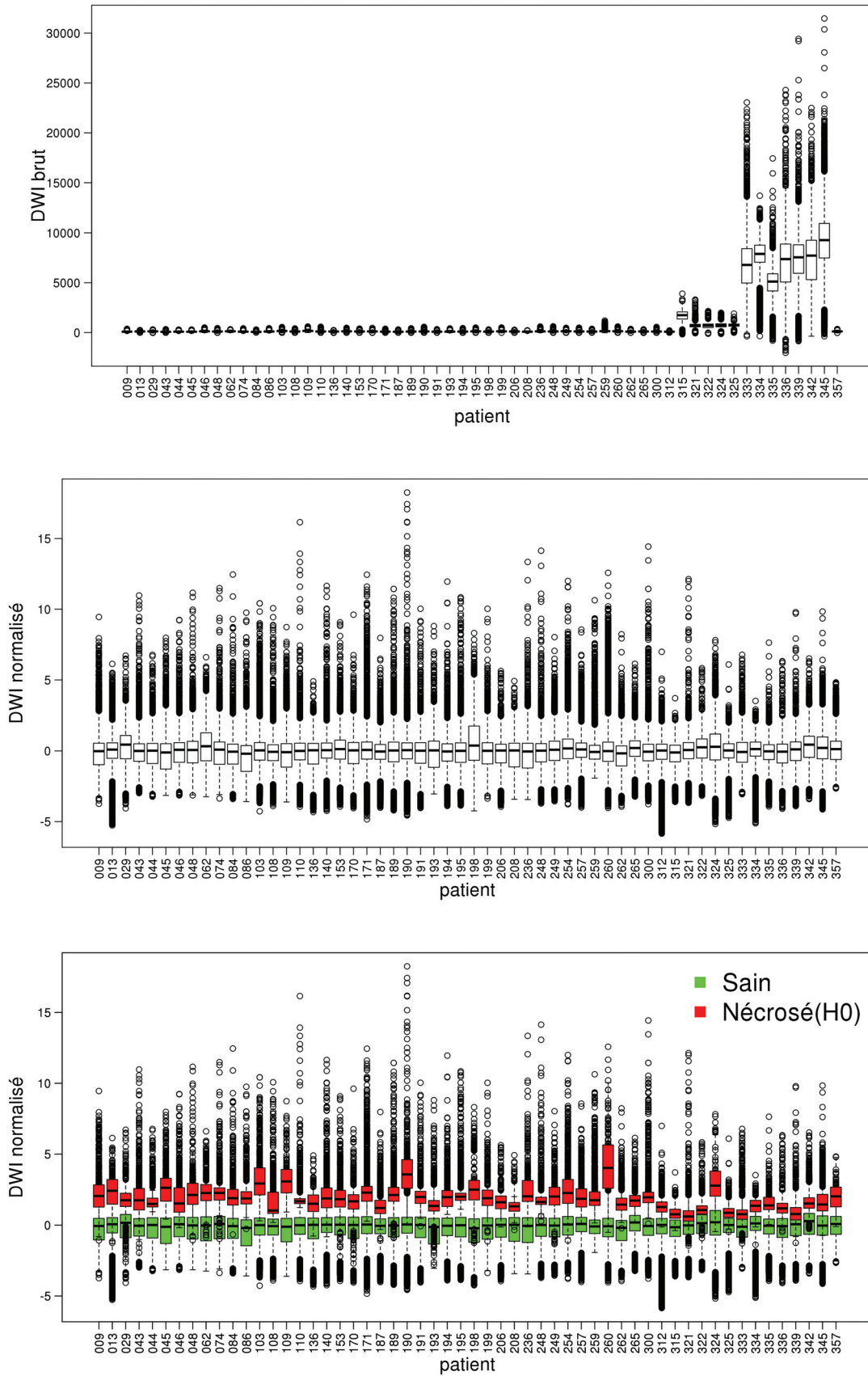


FIGURE II.7 – Distribution des valeurs de DWI par patient. En haut avant normalisation, au milieu et en bas après normalisation. Les graphiques du haut et du milieu représentent la distribution pour l'ensemble des voxels de chaque patient tandis que celui du bas différencie les voxels initialement nécrosés (en rouge) des voxels sains à l'admission (en vert).

3.3 Étude préliminaire : évaluation de l'intérêt des méthodes d'apprentissage automatique

L'objectif de cette section est d'évaluer l'intérêt des méthodes d'apprentissage automatique pour prédire l'évolution de l'AVC chez l'homme. Pour cela, quatre méthodes d'apprentissage automatique ont été comparées à la régression logistique : l'algorithme Adaboost, les réseaux de neurones artificiels, les forêts d'arbres décisionnels et les machines à vecteurs de support. Ce travail s'inscrit donc dans la continuité du travail de BOUTS et al., 2013, même si, en raison de la différence de nature des données, la méthodologie utilisée a été adaptée. Il a été réalisé par l'étudiante Céren Tozlu lors d'un stage de master 2 que j'ai co-encadré. Nous ne donnerons ici que les conclusions utiles pour la suite de cette thèse, un article reportant l'ensemble de l'étude étant en cours de rédaction.

La comparaison des modèles a mis en évidence une faible amélioration des performances prédictives voxel à voxel possible à l'aide de méthodes d'apprentissage automatique. L'algorithme Adaboost a montré les meilleures performances que ce soit en terme d'AUPRC, d'AUC ou de précision volumique, tandis que les machines à vecteurs de support montraient des performances plus modestes voire moins bonnes que le modèle logistique. Néanmoins les différences observées étaient très modestes (de l'ordre de +0,05 d'AUPRC), surtout comparées à la variabilité des performances selon les individus (AUPRC moyenne de 0,34 avec un écart-type de 0,27).

La prise en compte de la non-linéarité des fonctions de risque et d'interactions entre les variables prédictives ne semble donc pas réellement efficace. Cette absence de gain peut être imputable à la grande variabilité inter-individuelle qui rend l'estimation de fonctions de risque non linéaires plus difficile et limite son intérêt. La faible prévalence des voxels nécrosés pourrait aussi expliquer ce manque de performance : l'existence d'un déséquilibre important dans le jeu de données d'entraînement a été reporté comme néfaste pour les performances des approches prédictives (JONSDOTTIR et al., 2009). En effet les fonctions objectifs ne sont pas ajustées en fonction de la prévalence de chaque classe : la classe majoritaire a alors un poids prépondérant ce qui nuit à la qualité prédictive du modèle pour la classe minoritaire. Deux approches sont généralement employées pour faire face à ce problème (HE et al., 2009) : sous échantillonner le jeu de données d'entraînement afin d'obtenir un nombre équilibré de représentants de chaque classe ou pondérer les observations pour tenir compte de la prévalence dans la fonction objectif. Ces approches ont été testées sur un petit échantillon de patients ($n=5$). La pondération (par l'inverse de la prévalence de la classe) n'apportait pas de bénéfice net (amélioration ou dégradation des performances suivant le patient et le modèle) tandis que le rééchantillonnage nuisait aux performances. Ces corrections n'ont pas été retenues par la suite.

Comparaison avec le travail de Bouts et al., 2013

Même si nous avons souhaité poursuivre le travail de BOUTS et al., 2013, il y a des différences méthodologiques entre ce travail et le nôtre au niveau de la normalisation des données, de l'entraînement des modèles prédictifs, et des paramètres IRM utilisés. Néanmoins, nous avons trouvé des résultats concordants avec ceux de Bouts, même si les différences que nous avons obtenues étaient plus marquées entre les modèles : différence significative entre les performances de l'algorithme Adaboost et celles de la régression logistique après prise en compte de la variabilité patient. Cela peut être imputable au plus grand nombre d'individus dont nous disposions (n=55 comparé à 3 groupes de 14 dans Bouts).

3.4 Améliorations proposées

L'utilisation de méthodes d'apprentissage automatique ne suffit pas pour obtenir des prédictions fiables. Il est nécessaire de mieux prendre en compte les spécificités de l'AVC dans l'approche prédictive. Nous allons proposer dans cette partie trois voies d'amélioration qui seront évaluées par une étude comparative dans la section suivante.

Intégration de nouvelles variables prédictives

En plus des paramètres IRM précédemment utilisés, nous proposons d'inclure dans le modèle des variables spatiales et anatomiques :

- **distance** : la distance à la lésion initiale qui prend la valeur 0 pour les voxels inclus dans la lésion initiale et la distance au plus proche voxel de la lésion initiale pour les autres voxels. Cette variable permet de pénaliser les voxels éloignés de la lésion, et donc d'être plus robuste face aux artéfacts provenant des coupes basses. Près de la lésion, elle va permettre de tenir compte des réactions nocives générées par la dégradation des cellules voisines (inflammation, émission de radicaux libres ou de glutamate, MOSKOWITZ et al., 2010).
- **volume** : le volume de lésion initial permettant d'ajuster le modèle à la prévalence de la nécrose.
- **WM** et **GM** : le pourcentage en substance blanche et substance grise. Ces deux types de tissus ont des sensibilités différentes à l'ischémie (FALCAO et al., 2004; ARAKAWA et al., 2006; KOGA et al., 2005) et devraient donc être associées à des facteurs de risque différents.

Une des limites évidentes des variables spatiales (distance et volume) tient au fait que leur valeur ne reflète qu'un état transitoire (au moment de la première IRM du patient) tandis que toutes deux évoluent avec l'AVC. Pour cela, un modèle dynamique serait nécessaire afin de simuler au cours du temps l'évolution de l'AVC. Cependant, l'utilisation de ces modèles pour l'AVC est très récente avec un premier modèle proposé en 2013 (REKIK et al., 2013) et amélioré depuis (REKIK et al., 2014). Bien que prometteuses, ces approches sont complexes et ont besoin

de plusieurs temps de mesure pour s'ajuster (au moins trois). Cela n'est pas compatible avec l'utilisation d'un modèle prédictif en phase aigüe de l'AVC. Aussi nous nous limiterons à une approche statique.

Amélioration du rapport signal sur bruit des prédicteurs

Le second type d'améliorations envisagées est l'utilisation de méthodes de filtrage afin d'améliorer le rapport signal sur bruit des variables prédictives. Certains articles (BAGHER-EBADIAN et al., 2011 ; LU et al., 2005) mentionnent déjà une étape de filtrage des paramètres IRM. Nous nous intéresserons aux filtres de lissage (aussi appelés filtres passe-bas) : leur objectif est d'éliminer le bruit de l'image (supposé correspondre aux petites fluctuations) pour faire ressortir le signal (supposé correspondre aux tendances à grande échelle). Pour cela, différents filtres sont classiquement utilisés : filtre Gaussien, filtre médian, filtre min-max, filtre bilatéral. Nous retiendrons ici trois méthodes de filtrage qui avaient été présentées à la section 3.2 page 47.

Remise en cause des hypothèses du modèle logistique

Le troisième type d'améliorations proposées concerne la modélisation. Le modèle logistique apparaît faire consensus comme modèle de base. Néanmoins il repose sur un certain nombre d'hypothèses qui ne sont pas vérifiées dans le cadre de l'AVC :

- **hypothèse de linéarité du risque lié aux variables prédictives** : en effet, on suppose qu'un doublement de la valeur du DWI engendre un doublement du risque sur l'échelle logistique. Or, en ce qui concerne les neurones, il peut y avoir des valeurs seuils, c'est-à-dire des valeurs à partir desquelles l'évolution de la cellule va être profondément modifiée. Nous relâcherons cette hypothèse en utilisant un modèle additif généralisé qui permet de modéliser les risques associés aux valeurs de paramètres par une fonction souple et non d'imposer une forme linéaire (droite).
- **hypothèse d'indépendance des effets des variables prédictives** : les risques estimés sur les paramètres de diffusion sont supposés s'ajouter aux risques estimés sur les paramètres de perfusion. Or un voxel avec un hypersignal en diffusion devrait nécroser qu'il soit ou non anormal en perfusion. En revanche pour les voxels normaux en diffusion, les paramètres de perfusion sont considérés comme pronostiques de l'évolution du tissu. Étant donné le nombre important de paramètres du modèle ($K=10$), il n'était pas envisageable de considérer l'ensemble des interactions, ni de spécifier leur forme fonctionnelle. Un modèle très souple comme les modèles d'apprentissage automatique semble bien adapté pour intégrer des interactions entre paramètres. Nous utiliserons donc l'algorithme Adaboost puisqu'il a donné les résultats les plus prometteurs parmi les méthodes d'apprentissage automatique (voir section 3.3).

- **hypothèse d'indépendance spatiale** : l'évolution d'un voxel est supposée indépendante de celle de ses voisins. Or, la nécrose est un phénomène régional : bien souvent les voxels vont nécroser par zones dépendant d'une source d'irrigation sanguine commune. Même si la variable de distance à la lésion est censée intégrer la majorité de l'information spatiale, un modèle autologistique va être utilisé pour capter l'information spatiale résiduelle.
- **hypothèse d'indépendance intra-patient** : pour des valeurs de paramètres IRM prédictives égales, l'évolution du tissu cérébral est supposée indépendante du patient. Cette hypothèse contraste avec la très forte hétérogénéité du devenir des patients observée en clinique. Des patients apparemment semblables peuvent avoir des évolutions cliniques très différentes. Un modèle mixte avec des effets aléatoires sur les différents paramètres IRM sera utilisé pour identifier et estimer la variabilité inter-patients.

3.5 Étude comparative des différentes améliorations proposées

L'objectif de cette étude est d'évaluer la pertinence des améliorations afin de proposer une méthode de prédiction optimale. La méthodologie employée ainsi que les résultats obtenus sont présentés dans les deux sous-sections suivantes. L'article correspondant est intégré dans la troisième sous-section et est suivi de quelques éléments de perspective.

Méthode

L'étude est réalisée sur 45 des 63 patients de la cohorte I-know. Huit patients ont été exclus à cause de données manquantes ou d'absence de lésion visible en diffusion à l'admission du patient à l'hôpital. Dix autres patients ont été exclus en raison d'un mauvais recouvrement de la lésion initiale par la lésion finale ($<25\%$), probablement lié à un décalage spatial qui n'a pas pu être corrigé lors de l'étape de réalignement. Le détail des critères d'inclusion des patients ainsi que du protocole d'acquisition et de traitement des données est précisé dans les deux premiers paragraphes du matériel et méthodes de l'article (page 64).

Dix modèles prédictifs et quatre méthodes de filtrage sont comparés. Les modèles prédictifs utilisés sont présentés dans l'annexe B, aussi seule leur spécification pour le problème traité est précisée ici :

- **TTP, MTT et TMAX** : ces trois approches considèrent uniquement les cartographies de perfusion TTP, MTT ou TMAX pour prédire l'étendue de l'AVC final. Même si ces approches ne reposent pas explicitement sur un modèle prédictif, elles ont été retenues parce qu'elles sont couramment utilisées dans la littérature (voir par exemple l'article de CHRISTENSEN et al., 2009).

- **LR** : modèle de régression logistique incluant l'ensemble des paramètres IRM. Il s'agit du modèle proposé dans WU et al., 2006, avec un paramètre IRM supplémentaire (le TTP). La probabilité de nécrose du voxel i du patient id est ajustée dans ce modèle selon la formule suivante :

$$\begin{aligned}
 \text{logit}(p_{i,id}) &= \alpha + \beta_{ADC} ADC_{i,id} + \beta_{DWI} DWI_{i,id} + \beta_{T2_{FLAIR}} T2_{FLAIR,i,id} \\
 &\quad + \beta_{CBF} CBF_{i,id} + \beta_{CBV} CBV_{i,id} \\
 &\quad + \beta_{TMAX} TMAX_{i,id} + \beta_{TTP} TTP_{i,id} + \beta_{MTT} MTT_{i,id} \\
 &= \alpha + \sum_{X \in \Omega} \beta_X X_{i,id}
 \end{aligned}$$

$$\text{avec } \Omega = \left\{ \begin{matrix} ADC, DWI, T2_{FLAIR} \\ CBF, CBV, TMAX, TTP, MTT \end{matrix} \right\}$$

- **LR+spatial** : modèle de régression logistique incluant à la fois les paramètres IRM ainsi que les paramètres spatiaux :

$$\begin{aligned}
 \text{logit}(p_{i,id}) &= \alpha + \sum_{X \in \Omega} \beta_X X_{i,id} \\
 &\quad + \beta_{distance} distance_{i,id} + \beta_{volume} volume_{i,id}
 \end{aligned}$$

- **LR+spatial+morpho** : modèle de régression logistique incluant à la fois les paramètres IRM ainsi que les paramètres spatiaux, en tenant compte d'une interaction avec le pourcentage de substance blanche et de substance grise pour l'effet des paramètres.

$$\begin{aligned}
 \text{logit}(p_{i,id}) &= \alpha + \alpha_{GM} GM_{i,id} + \alpha_{WM} WM_{i,id} \\
 &\quad + \sum_{X \in \Omega} \beta_X^{GM} X_{i,id} * GM_{i,id} + \beta_X^{WM} X_{i,id} * WM_{i,id} \\
 &\quad + \beta_{distance} distance_{i,id} + \beta_{volume} volume_{i,id}
 \end{aligned}$$

- **ADA+spatial** : algorithme Adaboost incluant les paramètres IRM ainsi que le volume et la distance à la lésion initiale. Le nombre d'arbres est cherché entre 1 et 150 et est déterminé en minimisant l'erreur de classification estimée sur les données non utilisées pour construire l'arbre ("out of bag error").
- **ALR+spatial** : modèle de régression logistique additif où le risque associé aux paramètres d'imagerie ainsi qu'à la variable de distance à la lésion est modélisé par des splines de régression plaques minces. Ces splines, notés s_X pour la variable prédictive X , sont entièrement déterminés par l'optimisation d'une vraisemblance pénalisée par un terme de régularité. Comme le suggère l'auteur de la méthode (S. N. WOOD, 2006,

section 4.5.3), le paramètre fixant le compromis entre l'ajustement aux données et la régularité des fonctions de risque est estimé par validation croisée généralisée.

$$\begin{aligned} \text{logit}(p_{i,id}) = & \alpha + \sum_{X \in \Omega} s_X(X_{i,id}) \\ & + \beta_{\text{distance}} \text{distance}_{i,id} + \beta_{\text{volume}} \text{volume}_{id} \end{aligned}$$

- **autoLR+spatial** : modèle de régression auto-logistique où la probabilité de nécrose d'un voxel dépend des probabilités de nécrose des voxels voisins à travers η , le paramètre de corrélation spatiale. La notion de voisinage est définie par la matrice de voisinage W qui a été calculée en ne considérant que les 10 plus proches voisins (8 dans le plan, un sur la coupe supérieure et un sur la coupe inférieure).

$$\begin{aligned} \text{logit}(p_{i,id}) = & \alpha + \sum_{X \in \Omega} \beta_X X_{i,id} \\ & + \beta_{\text{distance}} \text{distance}_{i,id} + \beta_{\text{volume}} \text{volume}_{id} + \eta W_{i,id} p_{id} \end{aligned}$$

- **MLR+spatial** : modèle de régression logistique à effets mixtes où un paramètre spécifique à chaque patient est estimé pour l'effet de chaque paramètre IRM ainsi que pour l'effet de la distance à la lésion initiale. La distribution des paramètres spécifiques aux patients est supposée suivre une loi normale.

$$\begin{aligned} \text{logit}(p_{i,id}) = & \alpha_{id} + \sum_{X \in \Omega} \beta_{X,id} X_{i,id} \\ & + \beta_{\text{distance},id} \text{distance}_{i,id} + \beta_{\text{volume}} \text{volume}_{id} \end{aligned}$$

Pour chaque approche prédictive quatre méthodes de filtrage des paramètres IRM sont considérées :

- **aucun filtrage**
- **lissage gaussien**
- **lissage bilatéral**
- **soustraction du controlatéral**

Chacun de ces filtrages a été présenté à la section 3.2 page 47. La normalisation des paramètres ainsi obtenus est faite comme décrit dans le tableau II.5 page 48. Pour limiter le temps de calcul, chaque méthode est estimée sur l'ensemble des patients mais uniquement sur les voxels de l'hémisphère touché par l'AVC et situés au plus à une coupe d'un voxel de la lésion finale ou initiale.

Les performances des différentes méthodes sont comparées par rapport à la valeur d'AUPRC. Les valeurs d'AUC et de précision volumique ($|\hat{V} - V|$) sont également reportées. La prédiction

du risque de nécrose avec le modèle logistique à effets mixtes est réalisée en ne considérant que la loi marginale. Par ailleurs, la possibilité de prédire les effets aléatoires à l'aide des variables cliniques : âge, score NIHSS initial, traitement, délai depuis les symptômes, taille de la lésion initiale, taille du vaisseau occlus (artère importante ou non) est évaluée en utilisant une régression linéaire multivariée. Une élimination descendante variable par variable au seuil de p valeur de 0,1 est réalisée pour ne conserver que les variables cliniques pertinentes. Le volume de nécrose prédit est obtenu par seuillage des probabilités prédites par patient :

$$\widehat{V}_{id} = \sum_{i=1}^{n_{Id}} (p_{i,id} > \hat{\tau})$$

au seuil optimisant de l'erreur de prédiction volumique :

$$\hat{\tau} = \underset{\tau \in [-\infty; +\infty]}{\operatorname{argmin}} \left(\sum_{id=1}^{Id} \operatorname{median}(|\widehat{V}_{id} - V_{id}|) + IQR(\widehat{V}_{id} - V_{id}) \right)$$

où *median* correspond à la médiane des valeurs et *IQR* à l'écart interquartile.

L'estimation des paramètres des modèles et du seuil pour le calcul des volumes est faite par validation croisée au niveau patient : le modèle est estimé sur l'ensemble des patients sauf un qui servira à évaluer le modèle. Le processus est répété autant de fois qu'il y a de patients. Le modèle optimal est déterminé en utilisant une analyse de variance (ANOVA) à trois facteurs sur les performances : un facteur modèle, un facteur filtrage et un facteur patient. Les comparaisons entre les performances des différents modèles ont été réalisées en corrigeant les p valeurs avec la méthode de Benjamini-Yekutieli afin de maintenir l'espérance de la proportion de faux positifs à moins de 5% (BENJAMINI et al., 2001).

Pour analyser l'impact d'une potentielle reperfusion, un seuil spécifique sera ajusté séparément chez les patients ayant reperfusés et chez les patients n'ayant pas reperfusés en utilisant comme définition de la reperfusion un critère en TTP précédemment publié (OZENNE et al., 2015). La différence de volume prédit servira à établir un critère de sélection des patients. Le critère d'éligibilité au traitement combinera deux aspects : absence de profil malin et récupération cliniquement significative si le traitement réussit à induire une reperfusion. Le profil malin sera défini par une lésion initiale de plus de 50 mL. Une récupération sera considérée comme significative si elle correspond à un volume d'au moins 6 mL et à au moins 25% de la lésion initiale. La différence d'évolution entre les patients traités validant le critère et ceux ne validant pas le critère sera évaluée au regard de la récupération relative du score NIHSS et de la croissance relative de la lésion.

Principaux résultats

L'ANOVA montre que le choix du modèle explique une part importante de la variabilité de l'AUPRC (28,2%) comparé au choix du filtrage (0,43%) même si la variabilité inter-patients reste prépondérante (71,3%). La régression logistique avec variables spatiales sur données lissées

apparaît comme le modèle le plus parcimonieux. Il est donc choisi comme modèle de référence. Le tableau II.6 détaille les performances des différentes approches prédictives. Le choix du filtrage apparaît avoir un effet marginal sur les performances (+0,047 d'AUPRC entre aucun filtrage et un lissage gaussien). La régression logistique multivariée est supérieure à l'utilisation d'un paramètre de perfusion seul (+0,089 d'AUPRC comparé au TTP seul). L'intégration des variables spatiales améliore plus nettement la capacité prédictive (+0,197 d'AUPRC) et améliore aussi l'ajustement du modèle (+15,9% de déviance expliquée). En revanche, l'inclusion des variables morphologiques ne semble pas être bénéfique ni pour l'ajustement (+0,1% de la déviance expliquée) ni pour la capacité prédictive du modèle (+0,003 d'AUC, p valeur de 1).

L'utilisation de modèles plus flexibles n'améliore pas significativement les performances. Le modèle le plus performant semble être la régression autologistique mais avec un gain en AUPRC faible (+0,012) et non significatif (p valeur de 1). Ces résultats sont confirmés par ceux obtenus pour l'AUC. Seul le modèle mixte apparaît offrir un bien meilleur ajustement aux données (+19,6% de déviance expliquée). Toutefois, le modèle marginal ne fournit pas des prédictions intéressantes (-0,338 d'AUPRC). En effet, l'écart type des effets aléatoires estimé par le modèle est important, généralement du même ordre de grandeur que l'estimation de l'effet fixe (voir le tableau 1 de l'article page 71). Les variables cliniques sont insuffisantes pour prédire ces effets aléatoires avec un coefficient de détermination R^2 ne dépassant pas 0,25 après élimination des variables non significativement associées. Le modèle retenu est donc le modèle logistique sur données lissées, avec variables spatiales mais sans variables morphologiques.

Les prédictions volumiques du modèle retenu ont une imprécision estimée à 7,86 mL. Cependant si l'on ignore l'effet patient, la variabilité de l'imprécision est très grande (écart interquartile de 21,91 mL). La corrélation entre le volume observé et le volume final est donc modeste ($\rho = 0,761$). L'estimation d'un seuil spécifique selon le statut de reperfusion aboutit à des seuils différents : 0,367 chez les reperfusés contre 0,534 chez les non-reperfusés. Les volumes prédits sur l'ensemble de la cohorte à partir de ces seuils sont aussi différents : l'écart minimal est de 0,94 mL, l'écart maximal de 47,9 mL et l'écart médian est de 7,32 mL (p valeur < 0,001). L'utilisation de ces volumes permet de sélectionner 19 patients : 14 ont été traités et 5 n'ont pas été traités. Les patients traités et sélectionnés sont comparables aux patients traités non sélectionnés en terme de gravité d'AVC (volume de lésion et score NIHSS). Parmi les patients traités, ceux sélectionnés montrent une diminution plus marquée de leur score de handicap NIHSS (-88,9 % contre -55,0%) même si la différence n'est pas significative (p=0,133). Cette différence entre sélectionnés et non sélectionnés est nettement moins marquée chez les patients non traités (-62,5% contre -58,8% avec une p valeur de 0,70 ; voir figure 2 de l'article page 75).

<i>approche de référence</i>	AUPRC			AUC			$ \hat{V} - V $ (mL)		
	estimation	écart-type	p valeur	estimation	écart-type	p valeur	estimation	écart-type	p valeur
	Δ AUPRC			Δ AUC			$\Delta \hat{V} - V $ (mL)		
LR + spatial avec lissage gaussien	0,623	0,0185		0,964	0,0117		7,86	3,78	
<i>choix du filtrage</i>	estimation	écart-type	p valeur	estimation	écart-type	p valeur	estimation	écart-type	p valeur
aucun	-0,047	0,0071	<0,001	-0,023	0,0045	<0,001	1,41	1,45	1
lissage bilatéral	-0,03	0,0071	<0,001	-0,013	0,0045	0,028	1,58	1,45	1
soustraction du controlatéral	-0,048	0,0071	<0,001	-0,022	0,0045	<0,001	0,67	1,45	1
<i>choix du modèle</i>	estimation	écart-type	p valeur	estimation	écart-type	p valeur	estimation	écart-type	p valeur
	Δ AUPRC			Δ AUC			$\Delta \hat{V} - V $ (mL)		
TTP	-0,286	0,0106	<0,001	-0,167	0,0067	<0,001	4,16	2,17	0,539
MTT	-0,36	0,0106	<0,001	-0,171	0,0067	<0,001	11,36	2,17	<0,001
TMAX	-0,343	0,0106	<0,001	-0,181	0,0067	<0,001	3,83	2,17	0,684
LR	-0,197	0,0106	<0,001	-0,089	0,0067	<0,001	3,69	2,17	0,732
LR + spatial + morpho	0,003	0,0106	1	-0,001	0,0067	1	0,44	2,17	1
ADA + spatial	-0,004	0,0106	1	-0,002	0,0067	1	4,7	2,17	0,436
GAM + spatial	-0,005	0,0106	1	-0,002	0,0067	1	0,32	2,17	1
autoLR + spatial	0,012	0,0106	1	0,001	0,0067	1	0,26	2,17	1
MLR + spatial	-0,339	0,0173	<0,001	-0,184	0,0109	<0,001	25,46	3,54	<0,001

TABLE II.6 – Résultats de l'ANOVA sur les performances du modèle prédictif avec un effet patient, un effet modèle et un effet filtrage. 1^{er} tableau : performances de l'approche de référence. 2nd et 3^{ème} tableaux : influence du modèle et du filtrage sur les performances relativement à l'approche de référence. Les p valeurs sont ajustées par la méthode de Benjamini-Yekutieli.

Article

Le développement de modèles prédictifs pour l'AVC vise deux objectifs : mettre en place des critères d'éligibilité au traitement plus adapté que les critères actuels et permettre une meilleure évaluation des thérapies sur la base de critères d'imagerie. Comme rappelé précédemment, il s'agit de problématiques très proches de la pratique clinique : ainsi le modèle logistique proposé dans WU et al., 2006, a déjà été utilisé pour évaluer une oxygénothérapie à bas débit (WU et al., 2012). Des critères IRM de sélection de patients ont aussi été proposés et appliqués sur des cohortes de patients (ALBERS et al., 2006 ; LANSBERG et al., 2012). Aussi, il nous a semblé important de publier nos résultats dans une revue ayant un public à la fois médical et méthodologique. Néanmoins, cela suppose de simplifier le plus possible la méthodologie statistique pour rendre l'article lisible par un non-statisticien. Ainsi, la partie sélection de modèle a été réduite à l'essentiel tandis que la partie évaluation de l'intérêt clinique du modèle de prédiction a été développée.

Individualized Quantification of the Benefit from Reperfusion Therapy using Stroke Predictive Models

Brice Ozenne, MSc¹; Tae-Hee Cho, MD, PhD²; Irene Klærke Mikkelsen, PhD³; Marc Hermier, MD, PhD²; Lars Ribe, MSc³; Götz Thomalla, MD⁴; Salvador Pedraza, MD⁵; Pascal Roy, MD, PhD¹; Yves Berthezène, MD, PhD²; Norbert Nighoghossian, MD, PhD²; Leif Østergaard, MD, PhD³; Jean-Claude Baron, MD, PhD⁶; Delphine Maucort-Boulch, MD, PhD¹.

1. Service de Biostatistique, Hospices Civils de Lyon, Lyon, France ; Equipe Biostatistique Santé CNRS UMR 5558, Villeurbanne, France ; Université Lyon I, Lyon France.
2. Department of Stroke Medicine and Department of Neuroradiology; Université Lyon 1; CREATIS, CNRS UMR 5220-INSERM U1044 ; INSA-Lyon ; Hospices Civils de Lyon ; Lyon, France.
3. Center of Functionally Integrative Neuroscience, Århus University, Århus, Denmark.
4. Klinik und Poliklinik für Neurologie, Kopf- und Neurozentrum, Universitäts klinikum Hamburg-Eppendorf, Hamburg, Germany.
5. Department of Radiology (IDI), Girona Biomedical Research Institute (IDIBGI), Hospital Universitari de Girona DrJosep Trueta, Av de Francia s/n, 17007, Girona, Spain.
6. Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK ; Le Centre de Psychiatrie et Neurosciences, INSERM U894, Paris, France.

Abstract

Background and Purpose: Recent developments have shown the potential of voxel-based models in assessing the progression of stroke. Many models have been already proposed but their relevance in predicting the benefit of a reperfusion therapy remains unclear. We searched for an optimal predictive model whose volumic predictions would identify stroke patients likely to benefit from reperfusion.

Methods: Forty-five cases were used to study retrospectively stroke progression from admission to end of follow-up. Predictive approaches based on various statistical models, sets of predictive variables, and types of MRI parameter filtering were compared. Model accuracy was assessed using the area under the precision-recall curve (AUPRC) and the volumic error. Thresholds specific to the reperfusion status were used to compute volumic predictions from voxel predictions. The reperfusion-eligible profile was defined by an acute lesion of ≤ 50 mL and a difference between the predicted volume with and without reperfusion of at least 6 mL and $>25\%$ of the acute-lesion volume.

Results: The optimal model was a logistic regression using the voxel distance to the acute lesion, the volume of the acute lesion, and Gaussian-filtered MRI parameters as predictive variables. The predictions gave a median AUPRC of 0.655 and a median volumic error of 8.29mL. Nineteen patients matched the candidate profile. Reperfusion-eligible patients tended to have better functional recoveries than non-eligible patients (-88.9% vs. -55%, $p=0.13$).

Conclusion: Despite limited volumic accuracies, predictive stroke models can be used to quantify the benefit of reperfusion therapies and are thus suitable for deriving individual eligibility criteria.

Introduction

The clinical decision about thrombolysis treatment in stroke relies currently on a safety time window of 4.5 hours after stroke onset (Hacke et al. 2008). This has the immediate consequence of excluding nearly 20% of stroke patients with unknown time of symptom onset (Ona Wu, Schwamm, and Sorensen 2011). Additionally, a fixed safety time does not reflect the individual variability of stroke progress (Rekik et al. 2013). Then, more effective patient selection to treatment may be achieved considering individual stroke characteristics (Levine 2004).

The development of medical imaging in acute stroke has given a visual support to the follow-up of stroke progress. The diffusion sequence of magnetic resonance imaging (MRI) is currently the gold standard for the diagnosis of acute stroke despite discussions regarding its reliability in ischemic core identification (Kranz and Eastwood 2009). The present goal of several MRI studies is the estimation of the at-risk area (also called penumbra) because of its potential predictive value for treatment efficacy (Lansberg et al. 2012). During the last decade, the perfusion sequence, which aims to measure the hemodynamic compromise, has been extensively used to delineate the penumbra. However, no consensual criterion has emerged yet and the use of single parameters is considered too simple to capture the complex reality of stroke progression (Dani et al. 2011).

Recent approaches have promoted the use of multiparametric statistical models to better characterize the penumbra (L. Ostergaard, Jónsdóttir, and Mouridsen 2009). A variety of models have been proposed including logistic regression (Ona Wu et al. 2006) and machine learning methods (Bagher-Ebadian et al. 2011; Bouts et al. 2013; Huang, Shen, and Duong 2010, 2011). A few works have shown the usefulness of spatial information (Huang, Shen, and Duong 2010, 2011; Scalzo et al. 2012) or morphological variables (O Wu et al. 2004) but little attention has been given to the definition and the selection of the predictive variables to include in a prediction model.

Treatment allocation based on MRI profiles has been already studied (Albers et al. 2006; Lansberg et al. 2012); however, the proposed criteria were based on the extent of the

penumbra which is only indirectly linked to thrombolysis efficacy. Indeed, reperfusion therapies aim to induce a reperfusion of the penumbral tissue and enhance tissue recovery. The presence of a penumbra is thus a necessary but not sufficient indicator of the efficacy of reperfusion therapy.

We aimed to derive an individual quantification of the potential amount of tissue that would be saved by reperfusion therapy. We compared first several predictive approaches based on various statistical models, sets of predictive variables, and types of MRI parameter filtering in search for the optimal approach. We investigated then the way of including the treatment or the reperfusion status in the predictive approach to predict the potential benefit of reperfusion therapy.

Materials and Methods

Patients

We analyzed cases from the I-KNOW study (Ostergaard et al. 2006), a multicentric prospective study in which stroke patients underwent MRIs at admission, a few hours after admission, and at end of follow-up.

The inclusion criteria were: (1) a National Institutes of Health Stroke Scale (NIHSS) score ≥ 4 ; (2) diffusion-weighted and perfusion-weighted MR images consistent with an anterior-circulation acute ischemic stroke; (3) an admission MRI carried out within 6 hours in case of intravenous rt-PA (recombinant tissue-plasminogen activator) or within 12 hours in case of conservative treatment. The study excluded patients with lacunar or posterior circulation stroke, unknown time of symptom onset, intracerebral hemorrhage on MRI, missing MRI sequences, or lesion shifts on MRI between admission and follow-up. Finally, the present study kept 45 patients with complete imaging data.

Image acquisition and processing

At admission all patients underwent: a diffusion-weighted MR imaging providing an apparent diffusion coefficient (ADC) and a diffusion weighted imaging coefficient (DWI), a T2 fluid-attenuated inversion recovery (T2 FLAIR), a T2-weighted gradient echo, and a perfusion-weighted MR imaging. The latter technique provided a cerebral blood flow (CBF), a cerebral blood volume (CBV), a mean transit time (MTT), a time to maximum (TMAX), and a time to peak (TTP) for each voxel (For details about the acquisition and the signal processing protocols, see Ozenne et al. 2015). The reperfusion status of each patient (i.e., reperfused or non-reperfused) was assessed using a previously published criterion based on the percentage of reperfused tissue as measured by the TTP parameter at several thresholds (Ozenne et al. 2015).

MRI parameters could be not filtered, smoothed by applying a Gaussian or a bilateral Gaussian filter, or obtained by subtraction of the contralateral voxel contrast. A bandwidth of 1.875 mm was used for the Gaussian filters. Except when contralateral subtraction was performed, MRI parameters were normalized per patient using the mean and the standard

deviation of the contralateral tissues (cerebral spinal fluid excluded). All MRI parameters were scaled and centered to ensure comparability of the size effects.

Predictive approaches

The study compared four approaches able to predict the risk of infarction per voxel:

- Use of MTT, TMAX, or TTP perfusion parameter.
- A logistic model that included only MRI parameters: DWI, ADC, T2 FLAIR, CBF, CBV, MTT, TMAX, TTP as predictors of the risk of infarction. This approach is similar to that proposed by Wu (Wu O, 2006). It models the probability of infarction $P[Y_i = 1]$ by a linear risk function $risk(X_i) = \alpha + \sum_{j=1}^p \beta_j x_{i,j}$ on the logistic scale: $P[Y = 1] = \frac{1}{1+e^{-risk(X)}}$. Here, α denotes the baseline risk, p the number of predictors, $x_{i,j}$ the value of the j^{th} predictor for the i^{th} voxel, and β_j the risk factor associated with that voxel.
- A logistic model that includes the above-mentioned MRI parameters plus the distance of each voxel to the acute lesion and the volume of the acute lesion.
- More flexible predictive approaches that model a specific risk for each type of tissue, a non-linear risk function, or a spatial correlation. The use of machine learning methods was investigated through an adaptive boosting algorithm. More details about these approaches can be found in the online data supplement.

The optimal model was first identified then the ability of the treatment or the reperfusion variable to explain the remaining inter-patient heterogeneity was investigated by including, in the optimal model, interactions between each predictive variable and the treatment or the reperfusion variable.

Volumic prediction

A volumic prediction was obtained for each patient by thresholding the predicted probability of infarction:

$$\hat{V} = \sum_{i=1}^{n_{voxel}} (P[\text{infarction} = 1]_i > threshold)$$

The optimal infarction threshold was derived by minimizing the mean square volumic error:

$$\min_{threshold} IQR(\hat{V} - V_{final}) + median(|\hat{V} - V_{final}|)$$

where V_{final} denotes the final volume of the lesion and IQR the interquartile range.

Performance evaluation

Predictive approaches were assessed regarding the voxel and the volumic accuracy. Voxel accuracy reflects the ability of the predicted probabilities of infarction to distinguish between lesioned and unlesioned voxels. It was measured by the area under the precision-recall curve (AUPRC) which has been recommended in case of low prevalence (Ozenne, Subtil, and Maucourt-Boulch 2015). For comparison, the area under the ROC curve (AUC) was also reported. Volumic accuracy was measured by the absolute value of the difference between the predicted volume and the observed volume (absolute volumic error). The AUPRC, the AUC, and the absolute volumic error were determined in each patient using a leave-one-out cross-validation at the patient level: each patient is sequentially removed from the training set and used to assess the model performance.

Statistical Analysis

A three-way analysis of variance (ANOVA) was used to investigate the impact of the chosen statistical model and filtering method on the predictive performance after adjustment on the variability of that performance between patients. The optimal model performance with interaction was compared with its performance without interactions using Wilcoxon paired tests. The P values were adjusted for multiple testing using the correction of the false discovery rate (Benjamini and Yekutieli 2001). To assess inter-patient variability, the optimal model was fitted with patient-specific risk factors associated with each predictive variable, except the acute lesion volume. The standard errors of the patient specific risk factors were compared with the average risk factors.

The reperfusion-eligible profile was defined as that of a patient with a moderate acute lesion (≤ 50 mL) and a significant predicted benefit from reperfusion; i.e., a difference between the predicted volume with and without reperfusion of at least 6 mL and $>25\%$ of the acute lesion volume. Wilcoxon tests were used to compare the distributions of the NIHSS scores, the volumes of the acute lesions, and their progression between thrombolysis-eligible and not eligible patients and between reperfused and non-reperfused patients. Progression was defined by the ratio between the end of follow-up measurement minus the acute measurement, and the acute measurement.

Results

Comparison of the predictive approaches

Among the filtering methods, Gaussian filtering had the highest voxel accuracy ($p < 0.001$). However, it did not lead to a significant decrease of the absolute volumic error (-1.41 mL, $p = 1$). Single-parameter perfusion parameters were outperformed by the use of multivariable predictive models: TTP, which was found better than MTT or TMAX, had a lower voxel precision (-0.09 in the AUPRC, $p < 0.001$) despite a similar absolute volumic error ($+0.47$ mL, $p = 1$). The introduction into the model of voxel distance and volume as predictive variables led to a further improvement: $+0.20$ in the AUPRC ($p < 0.001$) and -3.69 mL in absolute volumic error ($p = 0.34$). The use of more flexible predictive models failed to improve significantly the voxel accuracy or the volumic accuracy. More details can be found in the online data supplement.

The retained approach was thus a logistic regression using voxel distance to the acute lesion, the acute lesion volume, and Gaussian filtered MRI parameters as predictive variables. The predictions gave a 0.655 median AUPRC (IQR: 0.260), a 0.976 median AUC (IQR: 0.032), and an 8.29 mL median absolute volumic error (IQR: 36.3 mL). Figure 1 shows a plot of the lesion volumes predicted by this approach vs. the observed final lesion volumes.

Analysis of the remaining inter-patient heterogeneity

The retained model explained only 45.27% of the deviance. Fitting a model with patient-specific risk factors led to explain a further 19.57% of the deviance: patient-specific risk factors showed a large inter-patient variability (see Table 1).

At hospital admission, 75% of the patients received a thrombolytic treatment. The AUPRCs were similar between treated and untreated patients ($p = 0.39$) but the absolute volumic error was higher in treated patients ($+4.94$ mL, $p = 0.02$). Adding interactions terms with the treatment status to the retained model brought a non-significant improvement of the voxel accuracy ($+0.00025$ in the AUPRC, $p = 1$) and no improvement of the absolute volumic error ($+0.19$ mL).

The clinical characteristics of the patients including the reperfusion status are summarized in Table 2. Twenty-five patients (55.6%) benefited from reperfusion. Compared with non-reperfused patients, they had greater reductions of NIHSS scores (-25.0%, $p=0.006$) and smaller lesion growths (-74.9%, $p=0.03$) but had initially lower NIHSS scores (-4.0, $p=0.025$) and smaller initial lesions (-27.7, $p<0.001$). Because large volumes are more prone to large errors than small ones, the absolute volumic error may be biased; the volumic error relative to the acute lesion volume is reported instead.

Table 3 displays the predictive performance of the optimal predictive model and of the TTP perfusion parameter according to the reperfusion subgroup. Voxel accuracy was significantly lower in reperfused patients with both approaches (e.g., 0.59 vs. 0.81, $p<0.001$ with the predictive model and AUPCR). Though the difference was not significant, the lesion volumes were overestimated in the reperfused group and underestimated in the non-reperfused group (e.g., +27.3% vs. -18.5, $p<0.52$ for the predictive model). Adding to the retained model interactions terms with the reperfusion status brought no improvement: -0.02 of AUPRC and +0.30 mL of absolute volumic error.

Defining eligibility criteria

The infarction thresholds were estimated at 0.53 in the reperfused group vs. 0.37 in the non-reperfused group. Using these thresholds for two volumic predictions (one for reperfusion and another for non-reperfusion) led to a difference in the predicted lesion volumes ranging from -0.94 mL to -47.9 mL, with a median of -7.32 mL. On the basis of this difference, 19 patients matched the reperfusion-eligible profile. Figure 2 displays the distribution of the acute NIHSS scores, the acute lesion, and their progression according to the profile. The 14 reperfusion-eligible patients who were treated had similar acute NIHSS scores ($p=0.78$) and similar acute lesion volumes ($p=0.76$) compared with the 19 non-eligible but treated. Both groups showed similar lesion growths ($p=0.78$) but the reperfusion-eligible patients tended to have a better decrease in NIHSS scores (-88.9% vs. -55%) despite a non-significant difference ($p=0.13$). Such a difference was not observed in the non-treated group.

Discussion

Using a multiparametric predictive model, we have shown the feasibility of an individualized quantification of the potential benefit of reperfusion therapy. By estimating the volume of tissue that would be saved in case of reperfusion, we enable the determination of clinically relevant eligibility criteria. Indeed, together, the potentially salvaged volume, the rate of reperfusion, and the hemorrhagic risk of the therapy help defining eligibility criteria that reflect the benefit-risk balance. In addition, our approach predicts an infarction risk and the salvaged volumes with and without reperfusion, which are valuable tools for decision making (see figure 3).

The currently available criteria are based on the extent of the penumbra: 10 mL and 120% of the acute lesion (Albers et al. 2006) or 15 mL and 180% of the acute lesion (Lansberg et al. 2012). However these criteria do not take into account of the heterogeneity of the penumbra: viability of the penumbral tissues depends of the degree and the duration of cerebral blood flow impairment (del Zoppo et al. 2011). A recent study (Kidwell, Wintermark, et al. 2013) has proposed the use of statistical modeling to distinguish salvageable from non-salvageable penumbral tissue. It stated that in stroke patients with recanalization (TIMI score ≥ 2), good functional outcomes were associated with moderate lesion sizes (acute lesion ≤ 90 mL) and limited non-salvageable penumbra ($\leq 70\%$). Unfortunately, this criterion was not able to identify stroke patients likely to benefit from endovascular therapy (Kidwell, Jahan, et al. 2013). This failure may be attributed to the inaccuracy of the model used to predict the percent of infarcted penumbra. Because of the very low prevalence of lesioned voxels, the use of the AUC (or percent of correct classification) may overestimate the model performance (Jonsdottir, Østergaard, and Mouridsen 2009; Ozenne, Subtil, and Maucort-Boulch 2015). Our predictive approach compares favorably with the one used by Kidwell et al. (Kidwell, Wintermark, et al. 2013) regarding voxel accuracy (0.98 vs. 0.76 AUC); besides, the latter approach showed a very moderate volumic accuracy with a large variability between patients. We think that the value of the AUPRC (0.655) would better reflect the true accuracy of our predictive approach.

The main improvement our approach brings is the inclusion of spatial descriptors (voxel distance to the lesion, volume of the lesion) as predictive variables of lesion extent. In comparison with the model proposed by Wu (Ona Wu et al. 2006), our model showed a

lower volumic error (6 ± 7 mL vs. 37 ± 35 mL in untreated patients and 29 ± 41 mL vs. 83 ± 55 mL in treated patients) and a higher AUC (0.98 ± 0.01 vs. 0.85 ± 0.06 in untreated patients and 0.96 ± 0.04 vs. 0.95 ± 0.05 in treated patients). Other predictive approaches with human data have shown close but inferior voxel accuracies: 0.91 AUC (Scalzo et al. 2012) and 0.89 AUC (Bagher-Ebadian et al. 2011). Machine learning methods were presented as promising tools for the prediction of stroke progression (Bagher-Ebadian et al. 2011; Huang, Shen, and Duong 2010, 2011). In one of our preliminary studies [*results not shown*], the adaptive boosting algorithm performed better than neural network, random forest, or support vector machines. However, the logistic regression was not found inferior to the adaptive boosting algorithm. This result is in line with the conclusion of Bouts on animal data (Bouts et al. 2013).

The present work has several limitations. First, the current model has an insufficient volumic accuracy for a clinical use. A reliable predictive model would need a volumic accuracy of only a few mL whereas the accuracy of the current model is about 10 mL. A large inter-patient variability remains and direct integration of clinical variables was not found relevant. A better use of clinical information, more advanced filtering methods (Forkert et al. 2013), and additional MRI parameters (Mouridsen et al. 2014) or spatial descriptors (Frindel et al. 2015) may improve the accuracy of our predictive model. However, unexpected sources of inaccuracy due to spontaneous reperfusions (Barber et al. 1998) or lesion atrophy (Ritzl et al. 2004) may persist. Second, because performance indices are not normally distributed, non-parametric tests (e.g., Wilcoxon test) may be more suitable than ANOVA. Nevertheless, ANOVA was still retained because it allows a concise presentation of the results, especially as Wilcoxon tests led to similar conclusions. Finally, we defined a reperfusion-eligible profile that seemed to identify patients with better functional recovery after thrombolysis compared with other patients. However, the I-know study was not designed to assess the merits of MRI-guided therapy. This profile needs thus confirmation and should be interpreted only as a proof of concept.

Table 1 - Coefficients estimated by the logistic mixed model.

Parameters	Risk factor estimate [*]	Standard error of the risk factor [*]	Standard error of the patient-specific risk factor [†]	P value [*]
Intercept	-2.40	0.46	2.31	<0.001
T2 fluid attenuated inversion recovery	0.66	0.14	0.88	<0.001
Apparent Diffusion Coefficient	-0.44	0.19	1.22	0.019
Diffusion Weighted Imaging	0.35	0.12	0.76	0.003
Cerebral Blood Flow	-0.27	0.19	1.2	0.156
Cerebral Blood Volume	-0.11	0.14	0.89	0.464
Mean Transit Time	0.16	0.11	0.71	0.155
Time to maximum	0.37	0.13	0.79	0.004
Time to peak	0.29	0.12	0.73	0.016
Voxel distance to the acute lesion	-4.42	0.33	2.16	<0.001
Volume of the acute lesion	1.01	0.38		0.008

^{*} Fixed-effect model- [†] Random-effect model.

Table 2 - Characteristics of the patients according to the reperfusion status.

Characteristics	Whole population	Reperused patients	Non-reperused patients	P value	Missing data
Number of patients	45 (100%)	25 (55.6%)	20 (44.4%)		
Age (yrs)	70	66	71	0.16	1
Females	20 (45.5%)	13 (54.2%)	7 (35%)	0.24	1
Acute NIHSS score	12	10	15	0.02	1
Time to MRI (h)	2.1	2	2.3	0.42	1
Acute lesion volume (mL)	18.2	8.5	42.9	<0.001	0
Etiology					
Large vessel disease	13 (28.9%)	8 (32%)	5 (25%)	0.36	0
Embolic stroke	17 (37.8%)	7 (28%)	10 (50%)	0.36	0
Thrombolysis	33 (75%)	18 (75%)	15 (75%)	1	1
NIHSS score progression (%)	-71.4	-85.7	-52.5	0.01	4
Lesion progression (%)	22.4	-43.6	42.2	0.03	1

Table 3 - Accuracy of the logistic model and the time to peak perfusion parameter in predicting the final stroke lesion according to the reperfusion status of the patients.

Indicator	Logistic regression			Time to peak		
	Reperfused	Non-reperfused	P value	Reperfused	Non-reperfused	P value
Median AUPRC (IQR)	0.59 (0.3)	0.81 (0.19)	<0.001	0.19 (0.41)	0.59 (0.39)	<0.001
Median AUC (IQR)	0.98 (0.03)	0.96 (0.03)	0.11	0.86 (0.15)	0.85 (0.09)	0.5
Median relative volumic error, % (IQR)	27.30 (202.15)	-18.54 (58.04)	0.52	25.29 (190.3)	-16.07 (74.37)	0.37

The relative volumic error is the difference between the predicted and the observed final lesion volume divided by the final lesion volume.

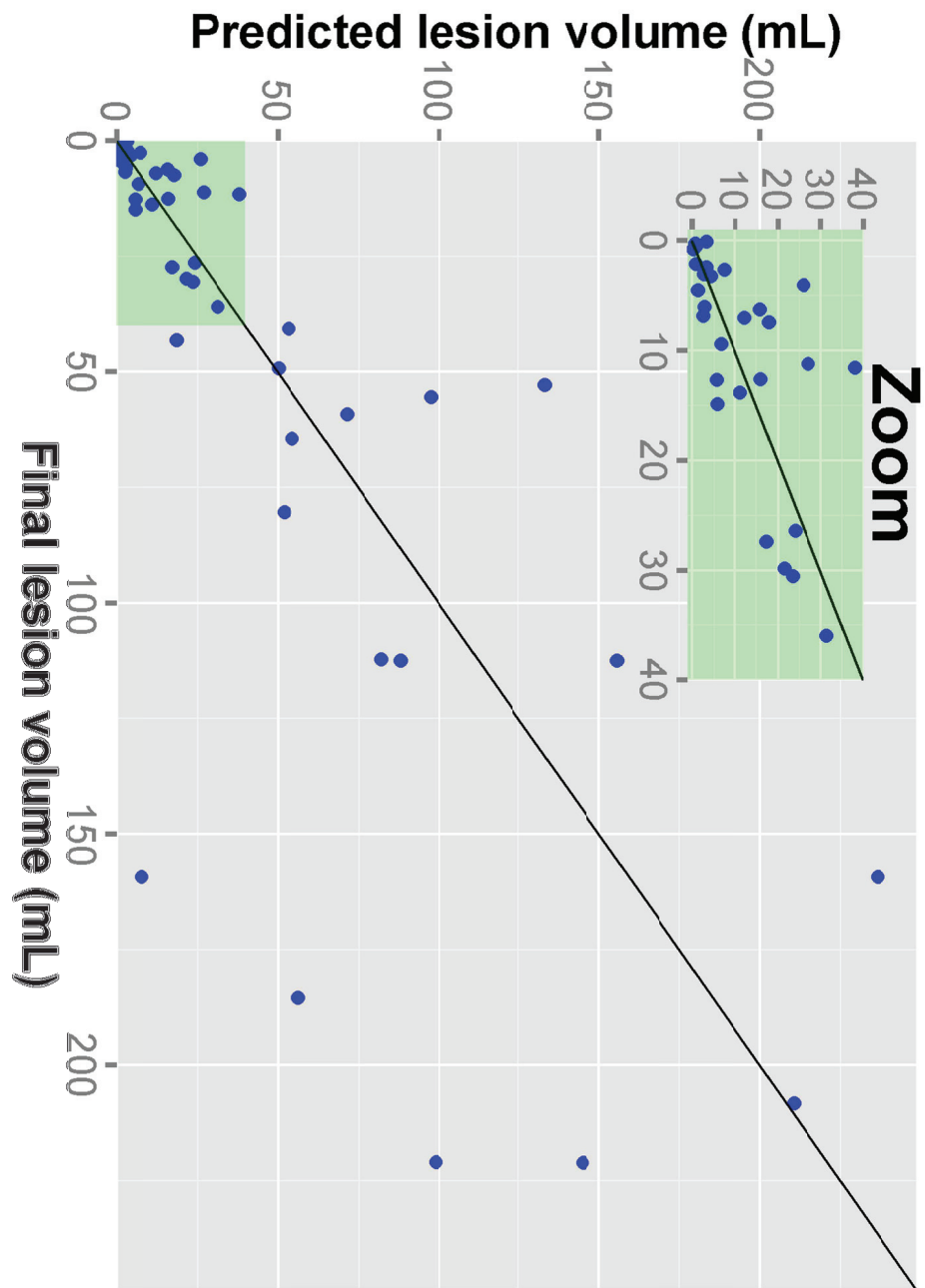


Figure 1 - Predicted vs. observed final lesion volumes.

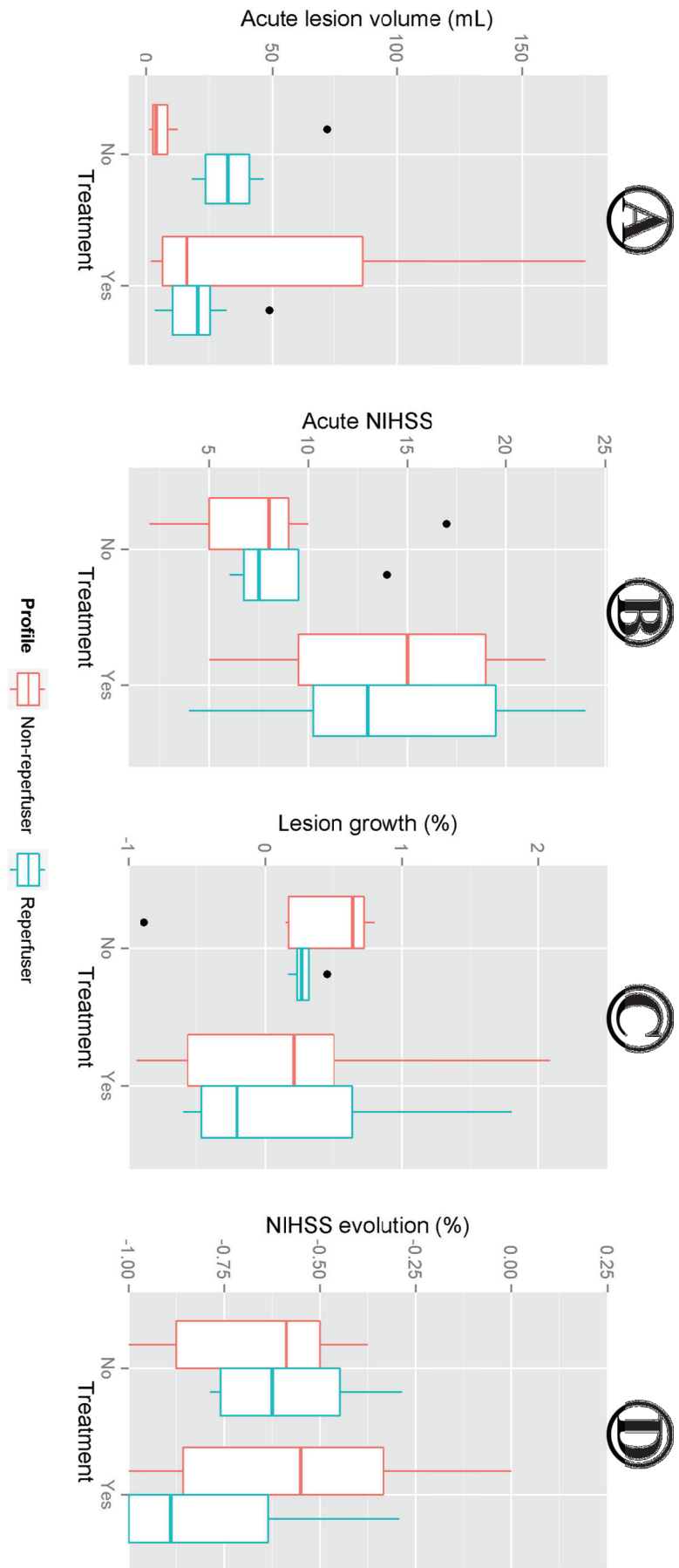


Figure 2 - Distribution of the acute lesion volume (A), the acute NIHSS score (B), the percentage of lesion growth between admission and end of follow-up (C), and the progression of the NIHSS score between admission and end of follow-up according to the treatment status (Yes/No) and the profile (left boxes for reperfusion-eligible patients and right boxes for non-reperfusion-eligible patients).

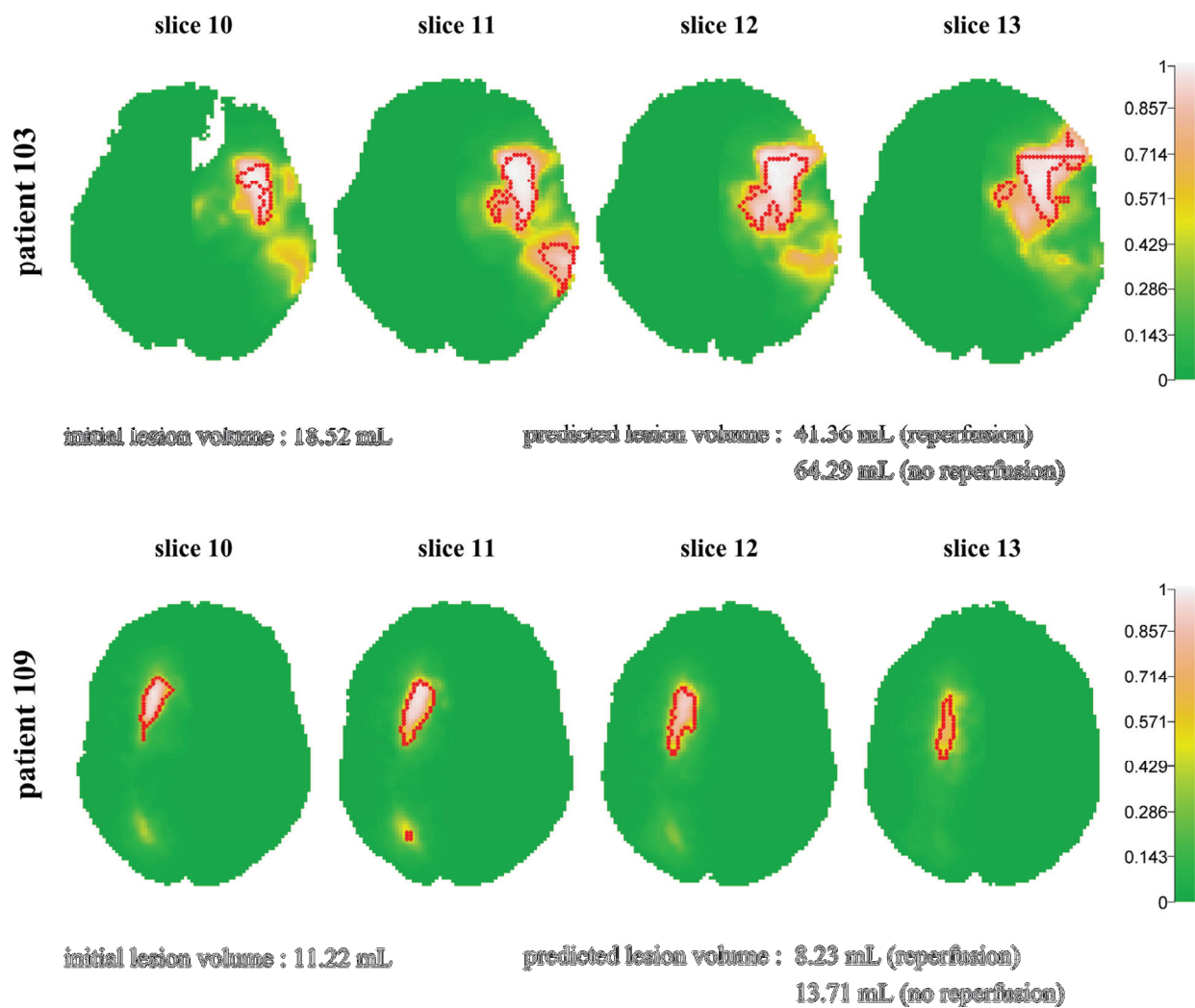


Figure 3 - Examples of a reperfusion-eligible (upper row) and a non-reperfusion-eligible profiles (lower row). The acute lesion is outlined in red. The volume of the acute lesion, the predicted volume of the final lesion in case of reperfusion and non-reperfusion are reported below the slices.

References

- Albers, Gregory W et al. 2006. "Magnetic Resonance Imaging Profiles Predict Clinical Response to Early Reperfusion: The Diffusion and Perfusion Imaging Evaluation for Understanding Stroke Evolution (DEFUSE) Study." *Annals of Neurology* 60(5): 508–17. <http://dx.doi.org/10.1002/ana.20976>.
- Bagher-Ebadian, Hassan et al. 2011. "Predicting Final Extent of Ischemic Infarction Using Artificial Neural Network Analysis of Multi-Parametric MRI in Patients with Stroke." *PLoS ONE* 6(8): e22626. <http://dx.plos.org/10.1371/journal.pone.0022626>.
- Barber, P Alan et al. 1998. "Spontaneous Reperfusion After Ischemic Stroke Is Associated With Improved Outcome." *Stroke* 29 (12): 2522–28. <http://stroke.ahajournals.org/content/29/12/2522.abstract>.
- Benjamini, Yoav, and Daniel Yekutieli. 2001. "The Control of the False Discovery Rate in Multiple Testing under Dependency." 29(4): 1165–88.
- Bouts, Mark J R J et al. 2013. "Early Identification of Potentially Salvageable Tissue with MRI-Based Predictive Algorithms after Experimental Ischemic Stroke." *Journal of cerebral blood flow and metabolism* □: official journal of the International Society of Cerebral Blood Flow and Metabolism 33(7): 1075–82. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3705436&tool=pmcentrez&rendertype=abstract>.
- Dani, Krishna a. et al. 2011. "Computed Tomography and Magnetic Resonance Perfusion Imaging in Ischemic Stroke: Definitions and Thresholds." *Annals of Neurology* 70(3): 384–401.
- Del Zoppo, Gregory J, Frank R Sharp, Wolf-Dieter Heiss, and Gregory W Albers. 2011. "Heterogeneity in the Penumbra." *J Cereb Blood Flow Metab* 31(9): 1836–51. <http://dx.doi.org/10.1038/jcbfm.2011.93>.
- Forkert, N D et al. 2013. "Comparison of 10 TTP and Tmax Estimation Techniques for MR Perfusion-Diffusion Mismatch Quantification in Acute Stroke." *American Journal of Neuroradiology* 34(9): 1697–1703. <http://www.ajnr.org/content/34/9/1697.abstract>.
- Frindel, Carole et al. 2015. "Validity of Shape as a Predictive Biomarker of Final Infarct Volume in Acute Ischemic Stroke." *Stroke*. <http://stroke.ahajournals.org/content/early/2015/03/05/STROKEAHA.114.008046.abstract>.
- Hacke, Werner et al. 2008. "Thrombolysis with Alteplase 3 to 4.5 Hours after Acute Ischemic Stroke." *New England Journal of Medicine* 359(13): 1317–29. <http://dx.doi.org/10.1056/NEJMoa0804656>.
- Huang, Shiliang, Qiang Shen, and Timothy Q Duong. 2010. "Artificial Neural Network Prediction of Ischemic Tissue Fate in Acute Stroke Imaging." *Journal of cerebral blood*

- flow and metabolism* □: official journal of the International Society of Cerebral Blood Flow and Metabolism 30(9): 1661–70. <http://dx.doi.org/10.1038/jcbfm.2010.56>.
- Huang, Shiliang, Qiang Shen, and Timothy Q. Duong. 2011. “Quantitative Prediction of Acute Ischemic Tissue Fate Using Support Vector Machine.” *Brain Research* 1405: 77–84. <http://dx.doi.org/10.1016/j.brainres.2011.05.066>.
- Jonsdottir, Kristjana Yr, Leif Østergaard, and Kim Mouridsen. 2009. “Predicting Tissue Outcome From Acute Stroke Magnetic Resonance Imaging: Improving Model Performance by Optimal Sampling of Training Data.” *Stroke* 40(1): 3006–11.
- Kidwell, Chelsea S, Reza Jahan, et al. 2013. “A Trial of Imaging Selection and Endovascular Treatment for Ischemic Stroke.” *New England Journal of Medicine* 368(10): 914–23. <http://dx.doi.org/10.1056/NEJMoa1212793>.
- Kidwell, Chelsea S., Max Wintermark, et al. 2013. “Multiparametric MRI and CT Models of Infarct Core and Favorable Penumbra Imaging Patterns in Acute Ischemic Stroke.” *Stroke* 44(1): 73–79.
- Kranz, Peter G., and J. D. Eastwood. 2009. “Does Diffusion-Weighted Imaging Represent the Ischemic Core? An Evidence-Based Systematic Review.” *American Journal of Neuroradiology* 30(6): 1206–12.
- Lansberg, Maarten G. et al. 2012. “MRI Profile and Response to Endovascular Reperfusion after Stroke (DEFUSE 2): A Prospective Cohort Study.” *The Lancet Neurology* 11(10): 860–67. [http://dx.doi.org/10.1016/S1474-4422\(12\)70203-X](http://dx.doi.org/10.1016/S1474-4422(12)70203-X).
- Levine, Steven R. 2004. “Optimizing an Individual’s Treatment in Acute Stroke: Is a Magnetic Resonance Map Leading Us towards the Holy Grail?” *Journal of the Neurological Sciences* 225(1-2): 1–2.
- Mouridsen, Kim, Mikkel Bo Hansen, Leif Ostergaard, and Sune Norhoj Jespersen. 2014. “Reliable Estimation of Capillary Transit Time Distributions Using DSC-MRI.” *J Cereb Blood Flow Metab* 34(9): 1511–21. <http://dx.doi.org/10.1038/jcbfm.2014.111>.
- Ostergaard, Leif (Coordinator) et al. 2006. “I-Know Integration Information from Molecule to Man.” *European Commission’s Sixth Framework Programme*. http://vph-portal.eu/documents/10180/35174/I_Know_factsheet.pdf.
- Ostergaard, Leif, Kristjana Ýr Jónsdóttir, and Kim Mouridsen. 2009. “Predicting Tissue Outcome in Stroke: New Approaches.” *Current Opinion in Neurology* 22(1): 54 – 59. <http://www.ncbi.nlm.nih.gov/pubmed/19155762>.
- Ozenne, Brice et al. 2015. “Evaluation of Early Reperfusion Criteria in Acute Ischemic Stroke.” *Journal of Neuroimaging* (2): n/a – n/a. <http://doi.wiley.com/10.1111/jon.12255>.
- Ozenne, Brice, Fabien Subtil, and Delphine Maucourt-Boulch. 2015. “The Precision Recall Curve Overcame the Optimism of the Receiver Operating Characteristic Curve in Rare

- Diseases.” *Journal of Clinical Epidemiology* 68(8): 855–59.
[http://www.jclinepi.com/article/S0895-4356\(15\)00106-7/abstract](http://www.jclinepi.com/article/S0895-4356(15)00106-7/abstract).
- Rekik, Islem, Stéphanie Allassonnière, Trevor K Carpenter, and Joanna M Wardlaw. 2013. “Spatiotemporal Dynamic Simulation of Acute Perfusion/Diffusion Ischemic Stroke Lesions Evolution□: A Pilot Study Derived from Longitudinal MR Patient Data.” *Computational and Mathematical Methods in Medicine*. <https://hal.inria.fr/hal-00935051>.
- Ritzl, Afra et al. 2004. “Development of Brain Infarct Volume as Assessed by Magnetic Resonance Imaging (MRI): Follow-up of Diffusion-Weighted MRI Lesions.” *Journal of Magnetic Resonance Imaging* 20(2): 201–7. <http://dx.doi.org/10.1002/jmri.20096>.
- Scalzo, Fabien et al. 2012. “Regional Prediction of Tissue Fate in Acute Ischemic Stroke.” *Annals of Biomedical Engineering* 40(10): 2177–87.
- Wu, O et al. 2004. “Anatomy as a Parameter in Multiparametric {MRI}-Based Predictive Algorithms.” *Proc. Intl. Soc. Mag. Reson. Med.* 11. <http://cds.ismrm.org/ismrm-2004/Files/001392.pdf>.
- Wu, Ona et al. 2006. “Characterizing Physiological Heterogeneity of Infarction Risk in Acute Human Ischaemic Stroke Using {MRI}.” *Brain* 129(9): 2384–93.
<http://www.ncbi.nlm.nih.gov/pubmed/16891322>.
- Wu, Ona, Lee H. Schwamm, and A. Gregory Sorensen. 2011. “Imaging Stroke Patients with Unclear Onset Times.” *Neuroimaging Clinics of North America* 21(2): 327–44.

Supplementary Data

Predictive approaches

We compared six predictive models and four image filtering methods to identify the optimal predictive approach. Two predictive models are presented in the main article: the logistic model inspired from Wu (Wu et al. 2006) denoted LR and the former model with two additional predictive variables (the voxel distance to the acute lesion and the volume of the acute lesion) denoted LR+spatial. The remaining predictive approaches are:

- LR+spatial+morpho: this is the same logistic model than LR+spatial, except that it contains interactions between the content in white and grey matter of each voxel and each predictive parameter (including the intercept). The contents in white and grey matter were estimated using a probabilistic segmentation algorithm (Feng and Tierney 2009) on the T1 contrast, when available or on the T2 weighted gradient echo contrast.
- ADA: this adaptive boosting algorithm is a machine learning method that aggregates sequentially several decision trees. Because decision trees partition recursively the data space with rectangular sets (i.e., $X \geq 2$ vs. $X < 2$, see (Loh 2011) for an introduction), they are suitable for modeling non-linearity and interactions between predictive parameters. The adaptive boosting algorithm improves the performance of a single decision tree by using additional trees to correct the classification errors made by the previous trees. The final classification is a sum of the predictions of each tree weighted by their error rates. We used the Discrete AdaBoost algorithm described in Friedman et al. (Friedman, Hastie, and Tibshirani 2000). The number of trees was set between 1 and 150 by optimizing the classification error.
- GAM: The generalized additive model extends the logistic regression by using a non-linear risk function $risk(X_i) = \alpha + \sum_{j=1}^p s_j(x_{i,j})$. Thus, a flexible risk function $s_j(x_{i,j})$ is adjusted to each parameter, for example, to model the risk maxima or the extra-risks linked with extreme values. Thin plate regression splines were used to model each risk function s_j with parameters estimated by maximizing a penalized log-likelihood. The

penalization parameter was estimated by minimizing the Generalized Cross-validation error. For a more comprehensive view see (Wood, 2009).

- autoLR: The autologistic model extends the logistic regression by modeling a dependency between the outcomes of neighbor observations. For this, it adds a spatial autocorrelation term to the risk function: $risk(X_i) = \alpha + \sum_{j=1}^p \beta_j x_{i,j} + \rho \sum_{j \in N(i)} Y_j$ where ρ denotes the autocorrelation parameter that controls the strength of the dependency and $N(i)$ the neighbor voxels of voxel i . Estimation was performed in two steps, first adjusting a logistic regression to estimate Y_j and then plugging-in this estimate and using the pseudo maximum likelihood approximation (Besag, 1975) to estimate ρ or predict $P[Y_i = 1]$.

These predictive models were compared with the use of TTP, MTT, and TMAX parameters, the common methodology to identify the at-risk tissue. Because cerebrospinal fluid voxels take often values close to those of lesioned voxels, they were removed to ensure a fair comparison with the predictive models.

Model training

To reduce the computational cost, the training of the models was restricted to the voxels that belong to a slice affected by stroke and the immediate upper and lower slice, excluding the voxels that belong to the cerebral spinal fluid.

Software

The analyses were carried out with R software (R Development Core Team, Vienna, Austria, <http://www.r-project.org>). The *MRlaggr* package was used for image processing and data normalization. The *stat* package was used for fitting the logistic regressions, the auto-logistic regression (under pseudo-maximum likelihood approximation), and the optimization of the infarction threshold. *mgcv* package was used for fitting the GAM model, *ada* package for the adaptive boosting algorithms, and *lme4* package for the mixed logistic model. R software and these packages are freely available at <http://cran.r-project.org/>.

Results

Table A summarizes the performance of each predictive approach according to the statistical model and the type of filter. LR + spatial was clearly superior to simpler models such as LR or the uniparametric approaches (TTP, MTT, TMAX) regarding all performance criteria. More flexible models led to similar performance. Because the latter models were more difficult to interpret and more demanding numerically (especially ADA and GAM), they were not considered in the rest of the study.

Table B displays the optimal infarction threshold that corresponds to each predictive approach. Considering non-filtered data, we found an optimal TMAX threshold of 6.44 s which is in the range of 6 to 8 s suggested by Shih et al. (Shih et al. 2003). This value was close to the threshold of 5.5 s found by Zaro-Weber et al. (Zaro-Weber et al. 2010) and to the range of 4 to 6 s suggested by Olivot et al. (J. M. Olivot et al. 2009) for defining the extent of the penumbra. The threshold of 11.52 s found for the MTT is higher than previously found thresholds : 7 s (Eilaghi et al. 2013) or 10 s (J.-M. Olivot et al. 2009). Regarding the predictive models, the optimal predicted probability of each voxel to belong to the lesion ranged from 0.37 to 0.49 indicating that voxels with moderate predicted probabilities of infarction are likely to belong to the lesion.

Table A - Results of the ANOVA.

Predictive approach	AUPRC* or Δ AUPRC			AUC* or Δ AUC			Absolute volumic error* or Δ AVE		
	Estimate	Std. error	P value	Estimate	Std. error	P value	Estimate	Std. error	P value
<i>Retained predictive approach</i>									
<i>Influence of the filter</i>									
LR + spatial with Gaussian filter	0.619*	0.0178*		0.960*	0.0107*		8.37*	3.16*	
<i>Influence of the model</i>									
No filter	-0.047	0.0067	<0.001	-0.023	0.0041	<0.001	1.41	1.19	1
Bilateral Gaussian filter	-0.030	0.0067	<0.001	-0.013	0.0041	0.009	1.58	1.19	1
Contralateral subtraction	-0.048	0.0067	<0.001	-0.022	0.0041	<0.001	0.67	1.19	1
<i>Influence of the model</i>									
TTP perfusion map	-0.286	0.0101	<0.001	-0.167	0.0061	<0.001	4.16	1.79	0.219
MTT perfusion map	-0.360	0.0101	<0.001	-0.171	0.0061	<0.001	11.36	1.79	<0.001
TMAX perfusion map	-0.343	0.0101	<0.001	-0.181	0.0061	<0.001	3.83	1.79	0.326
LR	-0.197	0.0101	<0.001	-0.089	0.0061	<0.001	3.69	1.79	0.341
LR + spatial + morpho	0.003	0.0101	1	-0.001	0.0061	1	0.44	1.79	1
ADA + spatial	-0.004	0.0101	1	-0.002	0.0061	1	4.70	1.79	0.133
GAM + spatial	-0.005	0.0101	1	-0.002	0.0061	1	0.32	1.79	1
autoLR + spatial	0.012	0.0101	1	0.001	0.0061	1	0.26	1.79	1

* Indices of accuracy of the retained predictive approach. Other values: differences in accuracy between the retained predictive approach and approaches using other filters or statistical models.

Table B - Median optimal threshold found with each statistical model and each type of filter after minimisation of the mean square volumic error in a leave-one-out cross-validation.

	No filter	Gaussian filter	Bilateral Gaussian filter	Contralateral subtraction
TTP perfusion map (s)	6.70	6.44	6.25	7.72
MTT perfusion map (s)	11.52	8.13	9.85	9.63
TMAX perfusion map (s)	6.44	4.32	5.37	6.44
LR	0.38	0.41	0.38	0.47
LR + spatial	0.45	0.43	0.44	0.43
LR + spatial + morpho	0.46	0.42	0.45	0.43
GAM + spatial	0.42	0.45	0.42	0.44
autoLR + spatial	0.43	0.45	0.45	0.43
ADA + spatial	0.38	0.49	0.43	0.39

References

- Eilaghi, Armin et al. 2013. "Reperfusion Is a Stronger Predictor of Good Clinical Outcome than Recanalization in Ischemic Stroke." *Radiology* 269(1): 240–48.
- Feng, Dai, and Luke Tierney. 2009. "MRI Tissue Classification Using High Resolution Bayesian Hidden Markov Normal Mixture Models." : 1–38.
- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2000. "Additive Logistic Regression: A Statistical View of Boosting." *Annals of Statistics* 28(2).
- Loh, Wei-Yin. 2011. "Classification and Regression Trees." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1(1): 14–23.
- Olivot, Jean Marc et al. 2009. "Optimal Tmax Threshold for Predicting Penumbra Tissue in Acute Stroke." *Stroke* 40(2): 469–75.
- Olivot, J-M et al. 2009. "Perfusion MRI (Tmax and MTT) Correlation with Xenon CT Cerebral Blood Flow in Stroke Patients." *Neurology* 72(13): 1140–45.
- Shih, Ludy C et al. 2003. "Perfusion-Weighted Magnetic Resonance Imaging Thresholds Identifying Core, Irreversibly Infarcted Tissue." *Stroke* 34(1): 1425–30.
- Wu, Ona et al. 2006. "Characterizing Physiological Heterogeneity of Infarction Risk in Acute Human Ischaemic Stroke Using {MRI}." *Brain* 129(9): 2384–93.
- Zaro-Weber, Olivier, Walter Moeller-Hartmann, Wolf-Dieter Heiss, and Jan Sobesky. 2010. "Maps of Time to Maximum and Time to Peak for Mismatch Definition in Clinical Stroke Studies Validated With Positron Emission Tomography." *Stroke* 41 (12): 2817–21.

Perspectives

L'étude a mis en évidence l'importance du choix des variables prédictives dans la modélisation. En effet, comme le montre le modèle à effets mixtes, il reste une grande hétérogénéité inter-patients qui n'est pas captée par les variables d'imagerie. Cette hétérogénéité provient en partie des profils cliniques différents des patients et aboutit à des différences de performance du modèle suivant la gravité de l'AVC (volume et score NIHSS), le délai depuis les symptômes et le traitement utilisé. Pour autant, ces variables n'apparaissent pas influencer directement la probabilité de nécrose de chaque voxel. Elles interviennent donc à un niveau plus global. Deux pistes ont été considérées pour prendre en compte les variables cliniques dans la modélisation : prédire les effets aléatoires associés à chaque patient ou ajuster la prédiction volumique sur les covariables cliniques (piste développée dans la section suivante). Néanmoins, les résultats obtenus n'ont pour le moment pas permis d'améliorer significativement la capacité prédictive du modèle.

Une limite importante de notre modélisation est l'utilisation d'un modèle statique tandis que l'AVC est un phénomène dynamique. Une modélisation spatio-temporelle pourrait permettre d'obtenir de meilleurs résultats : l'idée serait de modéliser l'AVC avec un maillage qui serait déformé en fonction de la cartographie des risques prédits et de la cinétique de l'AVC. Des modèles de déformation de formes ont déjà été formulés pour un paramètre de contraste (OSHER et al., 2003; ROSSO et al., 2009; REKIK et al., 2014). Leur extension à un contraste multiparamétrique étant difficile, notre approche peut être utilisée pour former un seul paramètre de contraste (le risque prédit) à partir des différents paramètres IRM. Néanmoins, une des difficultés sera d'estimer la cinétique de l'AVC, c'est-à-dire le temps d'expansion de la surface puisque celui-ci a été montré très variable entre les patients (REKIK et al., 2013).

Même si nous avons proposé un modèle prédictif permettant d'identifier les patients susceptibles de bien répondre au traitement, il souffre pour le moment d'une trop grande imprécision volumique pour être testé en clinique de façon similaire à l'étude DEFUSE (ALBERS et al., 2006). Une partie de cette imprécision est structurelle puisque lors de son évaluation, le modèle ne prenait pas en compte la survenue d'une reperfusion. Néanmoins cette variable n'explique pas complètement les erreurs de prédiction : après prise en compte du modèle, du filtrage et de la reperfusion il reste 73,17 % de la variabilité de l'AUPRC à expliquer. Même si, comme évoqué au paragraphe précédent, une meilleure prise en compte de l'information est souhaitable des progrès peuvent aussi être accompli au niveau de la mesure :

- la résolution IRM était ici limitée; une meilleure résolution spatiale améliorerait le contraste en limitant notamment l'effet de volume partiel.
- un meilleur traitement du signal (FRINDEL et al., 2015a) et de l'image permettrait aussi d'améliorer la qualité prédictive des paramètres de perfusion, notamment en retirant les artefacts près du liquide cérébro-spinal. Alors que des filtrage basiques ont été proposés, des améliorations ont déjà été observées. Ainsi, des méthodes plus avancées semblent avoir

tout leur place.

- de nouveaux paramètres IRM sont à l'étude comme la quantité d'oxygène extrait et l'hétérogénéité du temps de transit capillaire (MOURIDSEN et al., 2014). Ils permettraient d'augmenter l'information disponible conduisant ainsi à un modèle plus précis.
- l'identification du réseau vasculaire permettrait de définir une distance plus adaptée au problème que la distance euclidienne et donc une meilleure prise en compte de l'information spatiale.

L'ajout de variables prédictives, et plus particulièrement lorsqu'elles sont corrélées entre elles, pose le problème de la sélection des variables pertinentes. Cet aspect a été volontairement mis de côté dans l'étude. Nous aurions par exemple pu combiner les variables obtenues avec les différents filtrages et utiliser une procédure de sélection de variables. Dans une étude préliminaire, nous avons ainsi considéré deux types de variables pour chaque paramètre IRM : la valeur du voxel et la moyenne des valeurs des voxels voisins. Deux approches de sélection de variables avaient été utilisées : élimination pas à pas des variables non significatives après ajustement des effets aléatoires, ou une méthode LASSO pénalisant la vraisemblance afin de favoriser la parcimonie du modèle. Les deux méthodes aboutissaient à des jeux de variables proches : les paramètres IRM retenus étaient le TMAX, le DWI, le T2 et le CBF. Les performances entre les deux approches étaient similaires. Néanmoins, ces méthodes de sélection de variables souffrent d'instabilités et rendent les résultats peu lisibles. La préconisation de méthode de filtrage des paramètres IRM nous a semblé un message plus fiable et plus clair. Néanmoins, le problème risque à terme de se poser avec la multiplication des paramètres IRM.

3.6 Estimation du volume de nécrose à l'aide d'une approche fonctionnelle

La prédiction volumique à partir d'une cartographie de risque est un problème peu traité dans la littérature portant sur l'AVC. Seul un article a reporté l'erreur volumique associée aux volumes de lésion prédits (WU et al., 2006). Il s'agit pourtant d'un point de validation crucial du modèle qui pose pour le moment problème. En effet, l'article de WU et al., 2006, mentionne une erreur volumique de l'ordre de 37 ± 35 mL chez les patients non traités et de 83 ± 55 mL chez les patients traités. Sur nos données et avec le modèle proposé, nous trouvons une erreur certes moindre mais qui reste importante : 6 ± 7 mL chez les patients non traités et 29 ± 41 chez les patients traités. À noter que conformément à l'article de WU et al., 2006, nous reportons ici la moyenne \pm écart-type, même si celle-ci surévalue l'erreur en raison de valeurs extrêmes.

La méthode de prédiction volumique utilisée dans l'article de WU et al., 2006, a consisté à seuiller les risques prédits à 0,5 c'est-à-dire considérer que tout voxel ayant un risque au-dessus de 0,5 allait nécroser et tout voxel ayant un risque inférieur à 0,5 allait récupérer. Cette approche

est critiquable parce qu'elle n'interprète pas correctement les risques prédits par la régression logistique. En effet si le modèle est correctement calibré, une probabilité de nécrose estimée à 0,5 correspond à un voxel sur deux allant nécroser, tandis qu'une probabilité de nécrose estimée à 0,1 correspond à un voxel sur dix allant nécroser. Ainsi en sommant les risques, on obtiendrait une estimation non biaisée du volume de nécrose. Nous allons généraliser cette approche afin de pouvoir tenir compte de variables cliniques discrètes ou continues en se plaçant dans le cadre des modèles de régression statistique.

Principe de la méthode

Nous allons considérer le problème d'estimation du volume d'un point de vue fonctionnel (RAMSAY et al., 2005) : l'objectif est d'estimer la fonction g reliant le vecteur de probabilité de nécrose $(p_{i,id})_{i \in \{1, \dots, n_{id}\}}$ des voxels du patient id au volume de nécrose final V_{id} :

$$V_{id} = \sum_{i=1}^{n_{id}} g(p_{i,id}) + \varepsilon_{id}$$

avec ε_{id} le terme d'erreur volumique résiduelle. Dans notre application, la fonction g est supposée croissante et régulière. Cette approche généralise bien les méthodes de calcul de volume évoquées précédemment puisque l'on se ramène au seuillage en prenant g la fonction de Heavyside avec un saut en 0,5 et à la somme en prenant g la fonction identité (figure II.8).

La probabilité de nécrose étant une variable continue, l'estimation de la fonction g devient difficile. En effet, chaque prédiction $p_{i,id}$ étant considérée comme une variable prédictive, le nombre de variables devient grand devant le nombre de degrés de liberté du modèle. Afin de simplifier le problème on ne va plus considérer directement les prédictions mais l'histogramme des prédictions en associant à chaque classe de risque p^k un nombre de voxels n_{voxels}^k . On se ramène donc au problème suivant :

$$V_{id} = \sum_{k=1}^K n_{voxels,id}^k g(p^k) + \varepsilon_{id}$$

où K , le nombre de classes de risque, est pris égal à 50. Il s'agit d'un modèle additif généralisé avec un nombre fini de régresseurs p_{id}^k dont la seule particularité est que l'addition des effets des régresseurs se fait avec une pondération fixée a priori par le nombre de voxels correspondant à chaque classe de risques. Cette approche peut facilement inclure des covariables cliniques, notées \mathcal{C} , avec ou sans interaction avec la classe de risque :

$$V_{id} = \beta \mathcal{C}_{id} + \sum_{k=1}^K n_{voxels,id}^k g(p^k) + \sum_{k=1}^K n_{voxels,id}^k h(p^k, \mathcal{C}_{id}) + \varepsilon_{id}$$

Dans le cas d'une interaction avec les variables cliniques, le spline h est alors une fonction à plusieurs variables. L'estimation d'un spline multivarié est certes plus coûteuse mais tout à fait

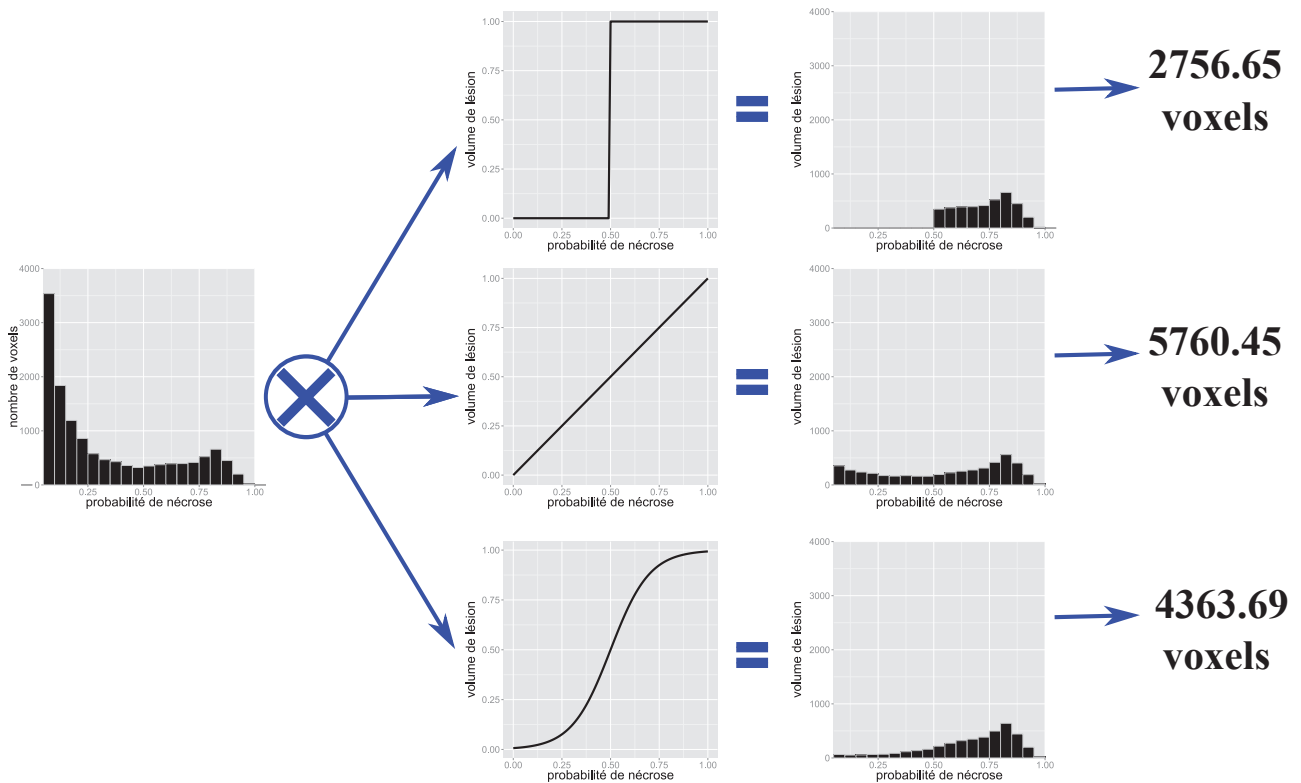


FIGURE II.8 – Principe de l’approche fonctionnelle pour l’estimation du volume à risque : l’histogramme des risques prédits est filtré par une fonction de risque (de haut en bas : fonction de Heaviside, fonction identité, fonction lisse) pour estimer le nombre de voxels qui vont réellement nécroser.

réalisable que ce soit avec des splines de régression ou avec des tenseurs (S. N. WOOD, 2006, section 4.1.8). L’utilisation d’une fonction lisse de type spline a deux intérêts : d’une part, elle permet d’imposer une fonction de risque régulière et, d’autre part, elle permet de régulariser le problème. En effet, le nombre de variables prédictives étant élevé (une par classe de risque de l’histogramme) par rapport au nombre de patients et ces variables étant très corrélées entre elles, la résolution du problème des moindres carrés associé souffre d’instabilités. En ajoutant un terme de pénalité sur la régularité de la fonction de risque, le problème devient mieux conditionné.

Objectifs

Cette nouvelle approche de prédiction volumique va nous permettre de tester :

- si le modèle logistique est correctement calibré, c’est-à-dire si les probabilités de nécrose reflètent bien les volumes finalement nécrosés.
- si l’utilisation d’une approche fonctionnelle permet une estimation plus précise des volumes de nécrose comparée au seuillage.
- si l’approche fonctionnelle permet de détecter l’effet des covariables cliniques dans l’évolution de l’AVC.

Méthode

La bonne calibration du modèle est évaluée en ajustant une fonction g plus souple que la fonction identité dans le calcul du volume :

$$g(p) = \alpha + (1 + \beta) \sum_{i=1}^n p_i + \gamma s(p)$$

Ainsi tester $\alpha = 0$ permet de vérifier l'absence de biais dans l'estimation des probabilités, $\beta = 0$ et $\gamma = 0$ l'absence de transformation, respectivement linéaire et non linéaire, des probabilités prédites qui améliorerait l'estimation du volume. La précision volumique du modèle fonctionnel retenu sera alors comparée à la précision volumique obtenue avec la méthode de seuillage. Un changement d'échelle effectué selon le paramètre estimé par la méthode de Box-Cox sera envisagé afin de stabiliser la variance des observations.

Ensuite, l'intérêt des covariables : âge, score NIHSS initial, traitement, délai depuis les symptômes, taille de la lésion initiale, taille du vaisseau occlus (artère importante : oui / non) pour la prédiction volumique sera testée. La prise en compte de la survenue d'une reperfusion sera envisagée afin d'estimer deux fonctions de risque permettant deux prédictions volumiques.

Résultats obtenus

La calibration du modèle est validée par l'approche fonctionnelle. L'écart à la linéarité est apparu non significatif (p valeur de 0,252). Le coefficient de pente estimé est très proche de 1 (différence de 0,06 avec une p valeur de 0,64) et le biais non significatif (-5,43 avec une p valeur de 0,39). La fonction somme non corrigée est donc utilisée pour la prédiction volumique. Les volumes prédits avec l'approche fonctionnelle montrent une corrélation quasi-identique à celle obtenue avec le seuillage (0,769 contre 0,761). Néanmoins, la médiane de l'erreur volumique absolue est plus élevée pour l'approche fonctionnelle (14,93 mL) que pour le seuillage (8,29 mL). De plus, l'analyse des résidus de l'approche fonctionnelle montre une tendance dans la variance des résidus. La méthode de Box-Cox propose une transformation racine quatrième. Après transformation, le coefficient γ reste non significatif (p valeur de 0,339) tandis que les coefficients β et α (p valeurs de 0,067 et 0,008) sont significatifs. Cette transformation permet de stabiliser la variance des résidus du modèle et de diminuer l'erreur volumique à 8,59 mL. Cette transformation est donc retenue pour la suite.

L'intégration des variables cliniques dans le modèle ne permet pas d'identifier d'influence significative. Le score NIHSS est cependant proche de la zone de significativité (p=0,10) et les volumes semblent être légèrement sous-évalués pour les score élevés ($\beta = 0,025$ par unité de score NIHSS sur l'échelle racine quatrième). L'inclusion du score apparaît aussi réduire l'erreur volumique à 7,89 mL, niveau proche de celui de la méthode de seuillage (p valeur de 0,31 ; voir figure II.9). La prise en compte dans le modèle de la variable reperfusion en interaction avec le risque de nécrose permet d'obtenir deux pentes : 0,95 en présence de reperfusion et 1,05 en l'absence de reperfusion. Néanmoins, ces deux pentes ne sont pas significativement différentes

(p valeur de 0,18). L'ajustement d'un spline sur les risques de nécrose, spécifique au statut de reperfusion, à la place d'une pente spécifique n'améliorait pas le modèle.

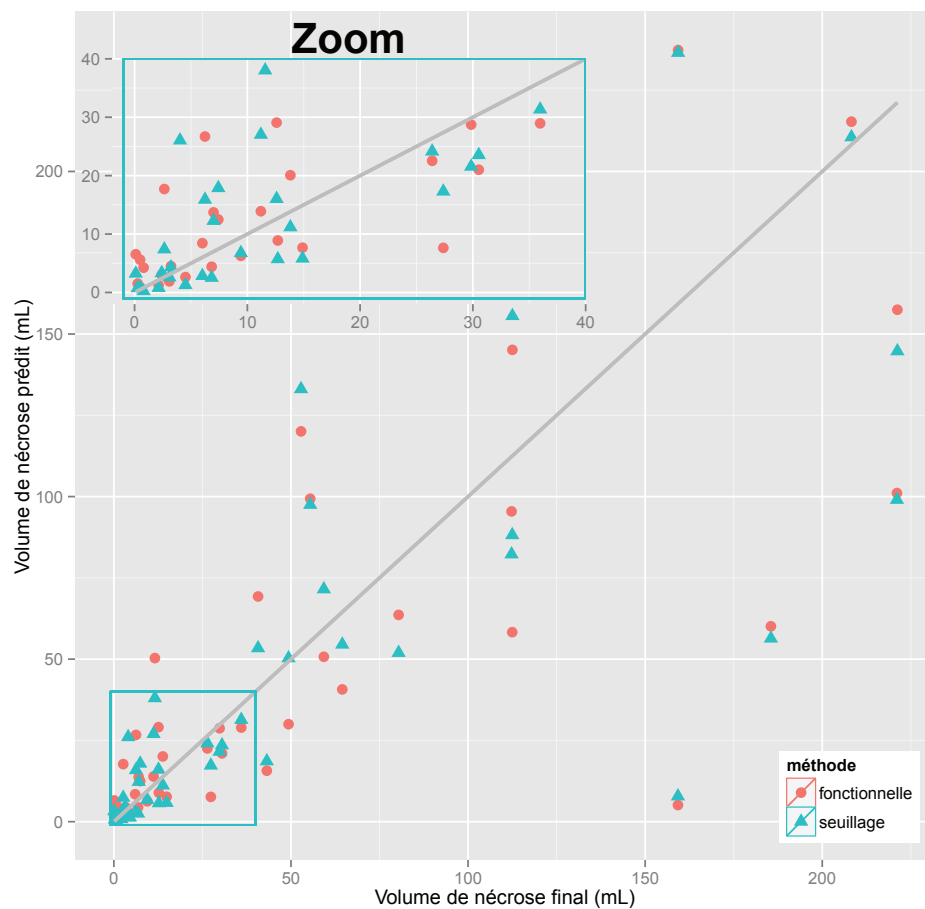


FIGURE II.9 – Comparaison entre les volumes prédits par l'approche fonctionnelle (disques) et la méthode de seuillage (triangles).

Discussion

L'approche fonctionnelle permet d'inclure des variables cliniques dans le modèle prédictif du volume de nécrose. En se replaçant dans le cadre statistique usuel via l'utilisation d'un modèle de régression, l'approche fonctionnelle permet d'évaluer l'influence des variables cliniques et éventuellement de stratifier la fonction de risque sur la variable de reperfusion. Elle peut donc être utilisée pour tester si un traitement est associé à une diminution du volume de nécrose : par exemple après ajustement sur les autres variables cliniques la thrombolyse était associée à un facteur protecteur ($\beta = -0,20$) mais non significatif (p valeur de 0,43). L'ajustement du modèle reste néanmoins délicat. Le volume est en effet une variable strictement positive dont la variance croît avec la valeur : l'hypothèse d'homoscédasticité n'est donc pas respectée. L'ajustement d'une régression gamma est conceptuellement plus satisfaisante mais posait des problèmes d'estimation. La transformation racine quatrième est apparue très satisfaisante en permettant d'être moins sensible aux erreurs correspondant aux AVC étendus.

L'approche fonctionnelle n'a néanmoins pas permis d'améliorer significativement la qualité de la prédiction volumique, ni d'identifier l'influence de la reperfusion. L'utilisation de fonctions très régulières (splines, effets linéaires) donne probablement trop d'importance aux risques faibles. Une future étude devra s'intéresser à des fonctions capables de modéliser des ruptures pour s'approcher de la fonction de Heaviside qui semble donner de meilleurs résultats.

4 Conclusion

Dans cette première partie, nous avons montré comment établir des critères d'éligibilité au traitement en utilisant les résultats d'un modèle de prédiction de nécrose. Le modèle prédictif voxel à voxel permet de former à partir de l'ensemble des données d'imagerie un biomarqueur reflétant le risque de nécrose de chaque voxel. Les valeurs de biomarqueur sont ensuite agrégées de façon propre à chaque scénario (absence ou présence de reperfusion) afin de proposer deux prédictions volumiques. La différence entre ces deux prédictions représente le gain potentiel du traitement et peut être mis en regard du risque hémorragique pour décider de thrombolyser ou non le patient.

Même si le modèle présenté montre des performances supérieures aux approches proposées dans la littérature, sa précision volumique reste toujours insuffisante. Cet aspect, peu mentionné dans les articles, peut être imputé à l'indice classiquement utilisé pour l'évaluation de modèles prédictifs (l'AUC) qui donne une vision optimiste des performances en imagerie. Nous avons donc proposé de substituer à cet indice l'AUPRC qui reflète mieux les performances réelles du modèle, ainsi que l'estimation de la précision volumique du modèle. En analysant l'apport de la modélisation, du filtrage des données et du choix des variables prédictives, nous avons obtenu un gain modeste apporté par les méthodes d'apprentissage automatique et par le filtrage tandis que l'inclusion de l'information spatiale s'est révélée déterminante dans l'amélioration des performances. Néanmoins de nouveaux paramètres IRM (MOURIDSEN et al., 2014), méthodes de traitement de signal (FORKERT et al., 2013) ou caractérisations spatiales de la lésion (FRINDEL et al., 2015b) pourraient améliorer la qualité du modèle. Enfin nous avons intégré la prédiction volumique au sein du cadre statistique usuel en utilisant une approche fonctionnelle. Cette approche permet d'intégrer les variables cliniques, jusque-là oubliées, dans la prédiction et d'évaluer l'impact du traitement sur le volume de lésion final.

Partie III

Développement d'outils d'analyse d'image

1 Segmentation multi-paramétrique de lésion avec régularisation de forme

1.1 Enjeu de la segmentation de lésion

Intérêt de la segmentation de lésion pour l'AVC

La segmentation de lésion est une problématique importante pour l'AVC mais aussi pour d'autres maladies comme la sclérose en plaque (LLADÓ et al., 2012), la maladie d'Alzheimer (LEVY-COOPERMAN et al., 2008), les tumeurs cérébrales (BAUER et al., 2013) ou encore la maladie de la substance blanche (YOSHITA et al., 2005). Dans le cadre de l'AVC, la segmentation de lésion sert plusieurs objectifs :

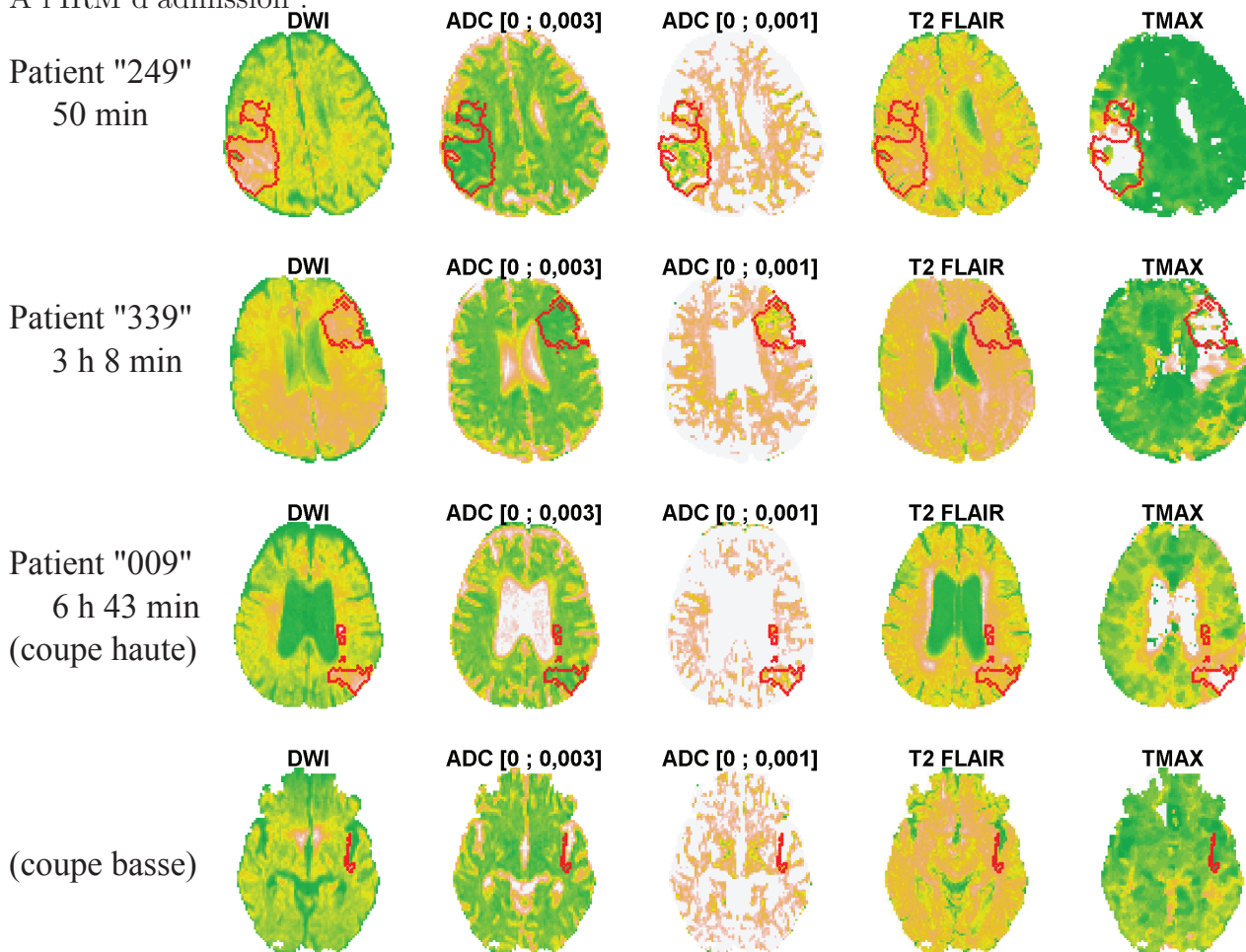
- **le diagnostic** : un diagnostic fiable d'un AVC ne peut être établi qu'à la suite de l'observation d'une lésion lors d'un examen d'imagerie médicale. S'il s'agit d'une IRM, le médecin est à la recherche d'une lésion en diffusion. Or, suivant le contraste de l'image, cette recherche peut être plus ou moins difficile : les images correspondant au patient "249" montrent un hypersignal en DWI (figure III.1, en brun ou blanc), un hyposignal en ADC (sous réserve de bien choisir l'échelle d'intensité) et une anomalie en TMAX. En revanche, les images correspondant au patient "339" montrent un hypersignal en diffusion peu marqué tandis que le TMAX permet de distinguer nettement la zone touchée par l'AVC. Une seconde information importante est l'étendue de l'AVC puisque l'on ne thrombolyse pas les patients ayant un AVC massif. En pratique clinique, l'étendue est estimée visuellement par un médecin. L'œil humain, surtout s'il est entraîné, est très efficace pour l'identification de lésions. Dans la majorité des cas, il est suffisant pour identifier efficacement les informations à extraire : présence ou absence, massif ou non massif. Néanmoins, lorsque le contraste est faible, une assistance informatique pourrait faciliter le travail du médecin.

- **l'évaluation de thérapies** : même si les essais cliniques utilisent les scores cliniques pour juger de l'efficacité du traitement, les articles associés reportent généralement les volumes de lésion (CLARK et al., 1999; CLARK et al., 2000; HACKE et al., 2009). D'autres études ont aussi évalué l'efficacité du traitement pour des patients ayant des profils spécifiques notamment en terme de volume de lésion : entre 10 et 100 mL (ALBERS et al., 2006), moins de 70 mL (LANSBERG et al., 2012). Cependant, le volume de lésion n'est pas systématiquement reporté même dans le cadre d'un essai clinique (THE NINDS GROUP, 1995; HACKE et al., 2008; LEES et al., 2010; IST-3, 2012).
- **l'ajustement de modèles prédictifs** : comme détaillé dans la partie précédente, les modèles prédictifs proposés dans la littérature sont principalement des modèles de prédiction supervisée. L'ajustement des paramètres du modèle se fait relativement aux lésions identifiées à l'IRM finale. Une mauvaise identification de la lésion finale entraînera un mauvais ajustement et donc de mauvaises prédictions.
- **l'étude de la cinétique de l'AVC** : des modèles spatio-temporels ont été proposés pour modéliser l'évolution des lésions en diffusion et en perfusion au cours du temps (REKIK et al., 2013; REKIK et al., 2014). Ces modèles visent à mieux comprendre la cinétique de l'AVC mais, comme les modèles de prédiction supervisée, leur validité repose sur une bonne identification de la lésion lors des différentes mesures IRM.

Difficultés spécifiques à la segmentation de lésion pour l'AVC

Comme l'illustre la figure III.1, les difficultés rencontrées pour la segmentation de lésion sont variables. Nous les classerons en trois catégories : (i) **contraste insuffisant** entre le tissu sain et la lésion (exemple : DWI patient "339", IRM d'admission), (ii) **hétérogénéité du signal** dans la lésion (exemple : ADC patient "249", IRM d'admission), (iii) **présence d'artéfacts** dus à la maladie de la substance blanche (exemple : T2 FLAIR patient "009", IRM final) ou provenant de structures cérébrales particulières (exemple : DWI patient "009", IRM d'admission, coupe basse). Par la suite, on entendra par artéfact tout hypersignal pouvant se confondre avec le signal de la lésion, mais ne correspondant pas à une zone touchée par l'AVC. La présence de nombreux artéfacts complexifie la reconnaissance de la lésion. Aussi, les algorithmes de segmentation sont généralement semi-automatiques : ils sont initialisés manuellement et proposent une segmentation qui est ensuite corrigée par le médecin. L'objectif est donc d'obtenir une segmentation qui soit robuste aux artéfacts et facile à initialiser afin de minimiser l'intervention du médecin.

À l'IRM d'admission :



À l'IRM finale :

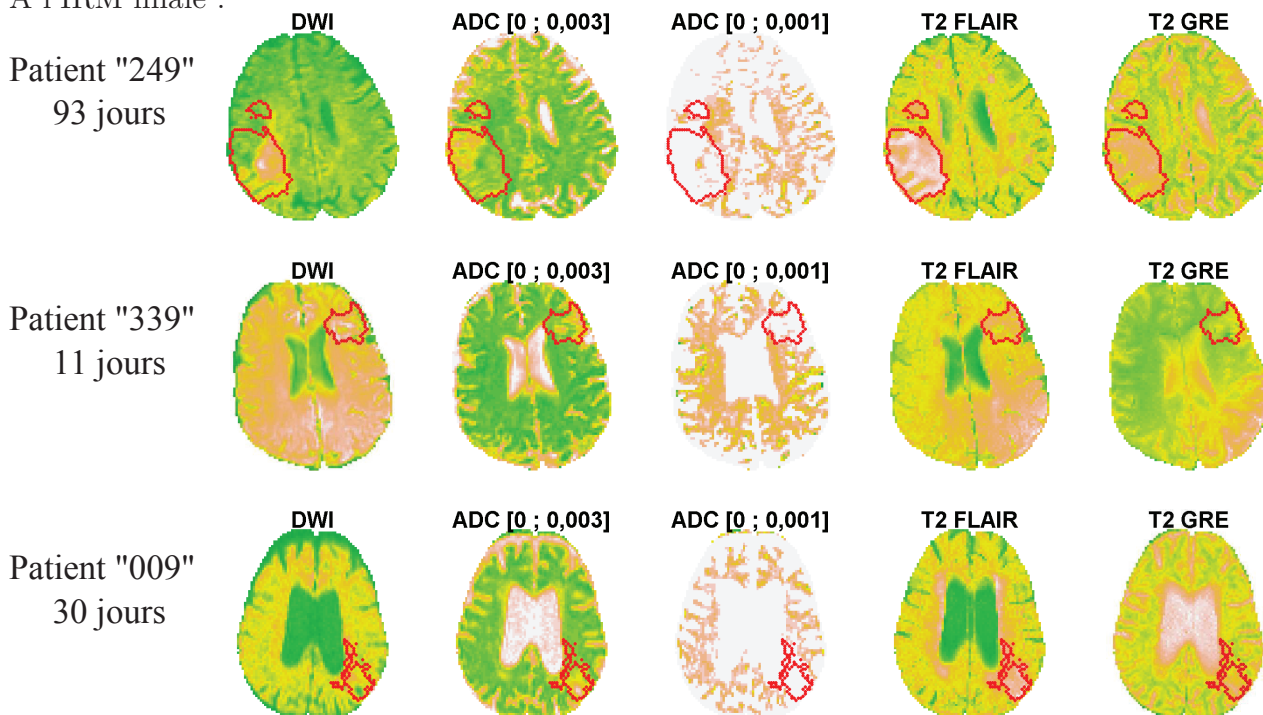


FIGURE III.1 – Exemple de cartographies IRM pour trois patients à l'IRM d'admission et à l'IRM finale. La lésion définie par le médecin est délimitée en rouge. Le temps indiqué correspond au délai entre l'examen IRM et le début des symptômes.

1.2 Les algorithmes de segmentation d'image pour l'AVC

Revue de la littérature

Le problème de segmentation d'image c'est à dire de partitionnement d'une image en régions, se pose dans de nombreux domaines : reconnaissance de visages (LI, 2009; VACHIER et al., 2005), restauration d'images (STORATH et al., 2014), identification de signal sur une puce ADN (GOTTARDO et al., 2006), analyse d'image aérienne ou satellite (PEREYRA et al., 2013), etc. Une littérature très vaste existe donc sur le sujet : voir par exemple l'article de DESPOTOVIĆ et al., 2015, pour une vue d'ensemble en imagerie cérébrale. Nous limiterons notre revue aux méthodes de segmentation ayant été évaluées sur des lésions d'AVC. D'une part, cela représente une littérature suffisante pour donner un bon aperçu des méthodes actuelles. D'autre part, comme nous l'avons vu au paragraphe précédent, certaines difficultés sont liées aux séquences d'imagerie utilisées pour l'AVC et nécessitent des approches spécifiques.

Une synthèse de la littérature existante sur les méthodes de segmentation de lésion pour l'AVC ayant été récemment publié (REKIK et al., 2012), nous reprendrons la bibliographie proposée en l'actualisant. Pour cela, une recherche sur les moteurs de recherche "Pubmed" et "Web of Science" a été effectuée avec les mots clés "stroke", "lesion", "segmentation". Même si les résultats obtenus ne sont pas exhaustifs, ils devraient être représentatifs de la littérature actuelle. Le tableau III.1 regroupe les différentes approches en cinq catégories :

- **les méthodes d'analyse d'intensité** : ces méthodes utilisent des heuristiques pour définir la lésion à partir des intensités observées. La méthode usuelle est le seuillage qui va identifier la lésion comme l'ensemble des voxels dont l'intensité excède un seuil donné. Ce seuil peut-être fixé au regard de la distribution d'intensité dans la zone saine (JAMES et al., 2006), ou estimé de façon adaptative en fonction de la segmentation obtenue (MARTEL et al., 1999). Afin d'améliorer la cohérence spatiale de la segmentation, MARTEL et al., 1999, utilisent un seuil spécifique à chaque voxel : la valeur du seuil dépend des groupes auxquels ont été associés les voxels voisins. Elle est estimée à l'aide de l'algorithme ICM (itération conditionnelle des modes). L'utilisation de plusieurs seuils a aussi été considérée. GHOSH et al., 2014, effectuent un seuillage itératif où, à chaque itération, le seuil d'intensité est déterminé par la méthode d'Otsu (minimisation de la variance intra classe). Une autre approche de partitionnement du volume cérébral en zones d'intensité homogène s'intéresse aux extrema locaux de l'histogramme d'intensité (BRAUN et al., 2002). Les minima locaux vont servir de seuils pour partitionner les voxels. Chaque partition a alors pour intensité, l'intensité du maximum local situé entre les deux minima locaux successifs. Enfin, USINSKAS et al., 2004, proposent d'utiliser, en plus de l'intensité, d'autres grandeurs caractéristiques comme la position dans la matrice de co-occurrence (mesure du type de texture de l'image) ou la position dans l'histogramme d'intensité.

Article	Nombre de patients	Paramètres d'imagerie	Type de méthode
Méthodes d'analyse d'intensité			
BRAUN et al., 2002	5 patients ‡	T2, DWI, ADC	Détection de groupes sur l'histogramme d'intensité
GHOSH et al., 2014	2 patients †	ADC	Seuillage itératif de l'intensité
JAMES et al., 2006	11 patients ‡	DWI, MTT	Seuillage
MARTEL et al., 1999	63 patients ‡	DWI	Seuillage + ICM
USINSKAS et al., 2004	non renseigné	CT	Combinaison de caractéristiques d'intensité
Méthodes d'apprentissage supervisé			
MAIER et al., 2015	37 patients	T1, T2w, T2 FLAIR	Combinaison d'arbres de décision (Extra Tree)
Méthodes de déformation de surfaces			
CONTIN et al., 2010	1 patient	CBF, CBV, MTT	Méthode de croissance de régions
HAAN et al., 2015	44 patients	CT, DWI, T2 FLAIR	Méthode de croissance de régions
GHOSH et al., 2014	2 patients †	ADC	Méthode de croissance de régions
HANSEN et al., 2015	167 patients	ADC, TTP	Méthode des surfaces de niveau
MATESIN et al., 2001	non renseigné	CT	Méthode de croissance de régions
MOURIDSEN et al., 2013	14 patients ‡	MTT	Méthode des surfaces de niveau
SAAD et al., 2012	20 patients	DWI	Croissance de régions
WEINMAN et al., 2003	5 patients	Non renseigné	Méthode des surfaces de niveau (diffusion non linéaire)

Article	Nombre de patients	Paramètres d'imagerie	Principe de la méthode
Méthodes de détection d'anomalie			
GILBERT et al., 2014	24 patients ‡	CT	Test de Crawford–Howell
MAH et al., 2014	435 patients	DWI	Zeta score (mesure d'anormalité)
MALDJIAN et al., 2001	15 patients	CT	Wilcoxon test
SEGHER et al., 2008	72 patients ‡	T1	Inconsistance de la classification avec des sujets contrôles
SHEN et al., 2008b	15 patients ‡	T1	Inconsistance entre un atlas et une classification floue
SHEN et al., 2010	29 patients ‡	T1	Inconsistance entre un atlas et une classification floue
WILKE et al., 2011	11 patients	T1	Z scores combinant 4 mesure d'anomalies
Méthodes de reconnaissance de groupes non supervisées			
ARTZI et al., 2013	44 patients	T2 FLAIR	MM avec régularisation spatiale
DALCA et al., 2014	106 patients	T2 FLAIR	MM avec a priori spatial
DWYER et al., 2008	5 patients ‡	ADC, MTT	HMMRF
FORBES et al., 2010b	4 patients ‡	T2, DWI, T2 FLAIR	MM bayésien avec régularisation spatiale
FORBES et al., 2010a	6 patients ‡	T2, DWI, T2 FLAIR	MM avec régularisation spatiale
JACOBS et al., 2001b	10 patients	T1, T2, DWI	Algorithme des k-moyennes modifié (ISODATA)
KABIR et al., 2007	56 patients	T2 FLAIR, DWI	MM avec régularisation spatiale
LI et al., 2004	20 patient ‡	DTI	MM avec régularisation spatiale
HEVIA-MONTIEL et al., 2007	15 patients	DWI	Algorithme 'Mean shift' pondéré
LI et al., 2009	19 patients	T2, DWI, ADC	Algorithme 'Mean shift'

Article	Nombre de patients	Paramètres d'imagerie	Principe de la méthode
Méthodes de reconnaissance de groupes non supervisées			
SOLTANIAN-ZADEH et al., 2007	15 patient ‡	T1, T2, DWI, PD	Algorithme des k-moyenne modifié (ISODATA)
Méthodes hybrides			
DASTIDAR et al., 2000	40 patients	T1, T2	Seuillage + Croissance de régions
MITRA et al., 2014	36 patients	FLAIR, ADC, T2, T1	MM + RF
PRAKASH et al., 2006	13 patients	DWI	ANN + MM
STEIN et al., 2001	6 patients	Non renseigné	Détection d'anomalie + surfaces de niveau

TABLE III.1 – Exemple de méthodes de segmentation de lésion ayant été évaluées sur des lésions d'AVC chez l'homme. L'algorithme ISODATA a également été évalué chez l'animal dans les travaux de JACOBS et al., 2000 ; JACOBS et al., 2001a ; SOLTANIAN-ZADEH et al., 2003.

- † : méthode évaluée sur données animales
- ‡ : méthode évaluée sur données simulées
- AMM : réseaux de neurones artificiels
- CT : tomographie calculée par ordinateur
- HMRf : modèle de markov caché
- ICM : itération conditionnelle des modes
- MM : modèle de mélange
- PD : densité protonique
- RF : forêts d'arbres décisionnels

- **les méthodes d'apprentissage supervisé** : cette classe de méthode a déjà été présentée dans la première partie du mémoire : à l'aide d'un jeu de données d'entraînement où l'on a identifié la lésion, ces méthodes vont chercher à apprendre comment associer les mesures IRM au statut nécrosé ou sain du voxel. Nous n'avons relevé que trois articles utilisant ce type d'approche pour segmenter une lésion d'AVC : les travaux de USINSKAS et al., 2003 ; PRAKASH et al., 2006, qui utilisent des réseaux de neurones artificiels et le travail de MAIER et al., 2015, qui utilise l'algorithme 'Extra-Tree'. Hors AVC, ces méthodes sont également utilisées avec par exemple les travaux de Sweeny et co-auteurs sur la sclérose en plaques (SWEENEY et al., 2013, SWEENEY et al., 2014) ou l'article de KLÉPPEL et al., 2011, portant sur l'identification de la maladie de la substance blanche.
- **les méthodes de déformation de surface** : ces méthodes vont partir d'une zone prédéfinie de l'image et modéliser sa frontière. A chaque pas, la frontière va être déformée pour étendre ou contracter la zone, afin de respecter un critère d'homogénéité (méthode de croissance de régions) ou de minimiser une fonction objectif (méthode des surfaces de niveau). Différents critères d'homogénéité ont été proposés pour les méthodes de croissance de régions : CONTIN et al., 2010, limitent les valeurs d'intensité possibles dans la région, HAAN et al., 2015, imposent un minimum d'intensité et met en compétition plusieurs régions (figure III.3 page 102). SAAD et al., 2012, limite la différence entre la valeur moyenne de l'intensité dans la région et l'intensité des nouveaux inclus dans la région. GHOSH et al., 2014, exploitent l'asymétrie de la distribution spatiale de l'intensité causée par la lésion : le critère d'homogénéité intègre alors l'asymétrie de la valeur d'intensité par rapport à l'hémisphère controlatéral et favorise l'appartenance à la lésion de voxels ayant une intensité asymétrique. À noter qu'il est possible de reformuler les méthodes de croissance de régions comme des méthodes de minimisation d'une fonction objectif (voir ROSE, 2008, chapitre 8). Nous avons trouvé deux implémentations de la méthode des surfaces de niveau dans la littérature. MOURIDSEN et al., 2013, utilisent une fonction objectif composée deux termes, un terme de fidélité favorisant l'homogénéité de l'intensité dans la zone de lésion, et un terme de régularité limitant la longueur du contour de la lésion. Cette approche a été évaluée sur large échantillon de patients dans HANSEN et al., 2015. Enfin WEINMAN et al., 2003, utilisent à la fois un critère d'intensité et de gradient d'intensité pour arrêter la propagation de la surface de niveau (figure III.2 page 102).
- **les méthodes de détection d'anomalie** : ces méthodes identifient la lésion comme des voxels anormaux par rapport à une référence. Ainsi GILLEBERT et al., 2014, utilisent le test de détection de données extrêmes de Crawford Howell pour comparer l'intensité observée en chaque voxel à la moyenne observées chez des patients contrôles. MAH et al., 2014, utilisent une procédure analogue avec un score d'anormalité appelé score zeta. WILKE et al., 2011, évaluent pour chaque voxel 4 mesures d'anormalité : homogénéité du contenu du voxel (en substance blanche, substance grise et liquide cérébro-spinal), concordance du contenu du voxel avec une population contrôle,

déterminant de la matrice jacobienne utilisée pour la normalisation spatiale, asymétrie de l'intensité et les combine avec un z score robuste. MALDJIAN et al., 2001, comparent la distribution régionale d'intensité entre l'hémisphère d'intérêt et l'hémisphère controlatérale avec un test de Wilcoxon. Shen et co-auteurs proposent une méthode comparant la classification floue des voxels du patient (fond, liquide cérébro-spinal, substance blanche et substance grise) à celle d'un atlas de référence. Deux métriques ont été proposées pour quantifier le degré de dissimilarité entre la classification et l'atlas (SHEN et al., 2008b ; SHEN et al., 2010). Enfin SEGHER et al., 2008, proposent de tenir compte dès la phase de classification de la présence de la lésion en ajoutant un groupe 'Extra' aux trois groupes usuels. L'identification de la lésion s'effectue en comparant la classification en substance blanche et substance grise des voxels du sujet à celle observée chez des patients contrôles (figure III.4 page 102). Une métrique de dissimilarité est utilisée pour fusionner le résultat obtenu pour la substance blanche et la substance grise.

- **les méthodes de reconnaissance de groupes non supervisées** : ces méthodes vont chercher à former des groupes de voxels en maximisant un critère d'homogénéité. De façon similaire aux méthodes de déformation de surfaces, ce critère est généralement composé d'un terme de fidélité aux données et d'un terme de régularité spatiale. Ces méthodes sont principalement issues des modèles de mélange qui modélisent l'intensité observée par un mélange de distributions. Ainsi, chaque observation a une probabilité d'appartenir à chaque groupe et chaque groupe a une distribution d'intensité spécifique. Le terme de fidélité correspond à la probabilité d'observer l'intensité, connaissant l'appartenance aux groupes et les paramètres de distribution des groupes. Le terme spatial est classiquement un critère d'homogénéité des groupes dans un voisinage (ARTZI et al., 2013 ; KABIR et al., 2007) : un voxel a davantage de chances d'appartenir à un groupe contenant déjà les voxels voisins. Ce terme spatial est fondé sur un modèle de markov caché (HMRF), qui est parfois directement utilisé pour effectuer la segmentation, comme par exemple dans l'article DWYER et al., 2008 (figure III.5 page 103). Les modèles de mélange offrent une grande souplesse au niveau la modélisation : ajout de termes corrigeant les artéfacts de segmentation causés par les fibres nerveuses (LI et al., 2004) ou par la maladie de la substance blanche (DALCA et al., 2014), modulation de l'importance des paramètres d'imagerie lors d'une segmentation multiparamétrique (FORBES et al., 2010b ; FORBES et al., 2010a). Des approches non-probabilistes ont aussi été proposées : JACOBS et al., 2001b, et SOLTANIAN-ZADEH et al., 2007, ont utilisé l'algorithme ISODATA, un algorithme des k-moyennes modifié, pour évaluer l'étendue de la zone de pénombre (figure III.6 page 103). Enfin l'algorithme 'Mean Shift' a été proposé comme alternative aux approches précédentes (LI, 2009 ; HEVIA-MONTIEL et al., 2007). Cette méthode est à la fois non paramétrique et sans a priori sur le nombre de groupes à identifier (figure III.7 page 103).

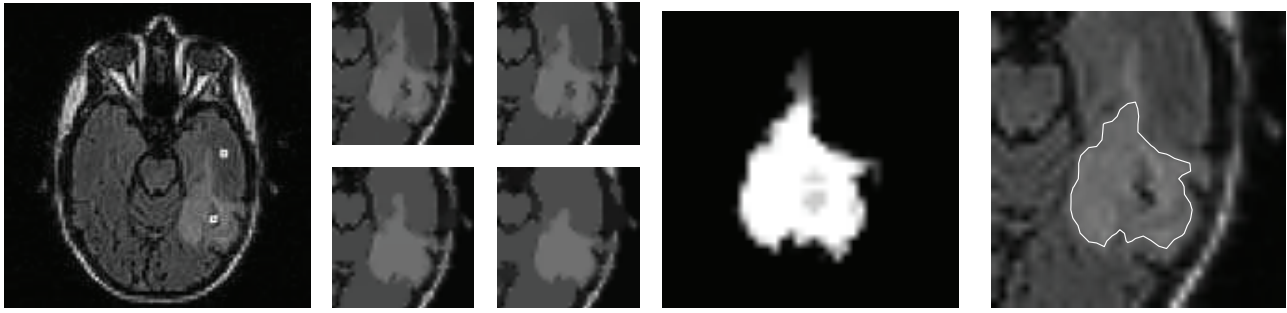


FIGURE III.2 – Application d'une méthode de surface de niveaux en segmentation de lésion (WEINMAN et al., 2003). De gauche à droite : initialisation de la zone saine et de la zone nécrosée, résultats de l'algorithme avec 4 paramètres d'échelle différents, zone de segmentation commune et contour obtenu pour une confiance de 0,71.

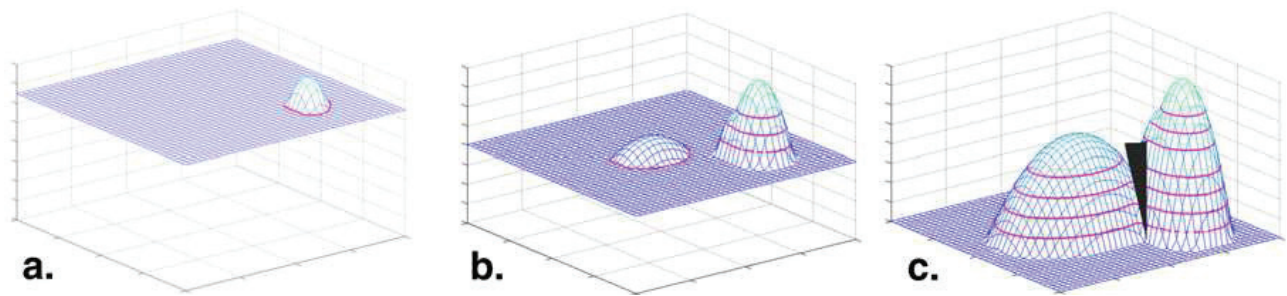


FIGURE III.3 – Algorithme de croissance de régions utilisé par HAAN et al., 2015. L'algorithme est décrit dans l'article de CLAS et al., 2012, dont est issu la figure. Deux régions sont mises en compétition. Par pas d'intensité de 1%, les voxels sont affectés à l'une des deux régions.

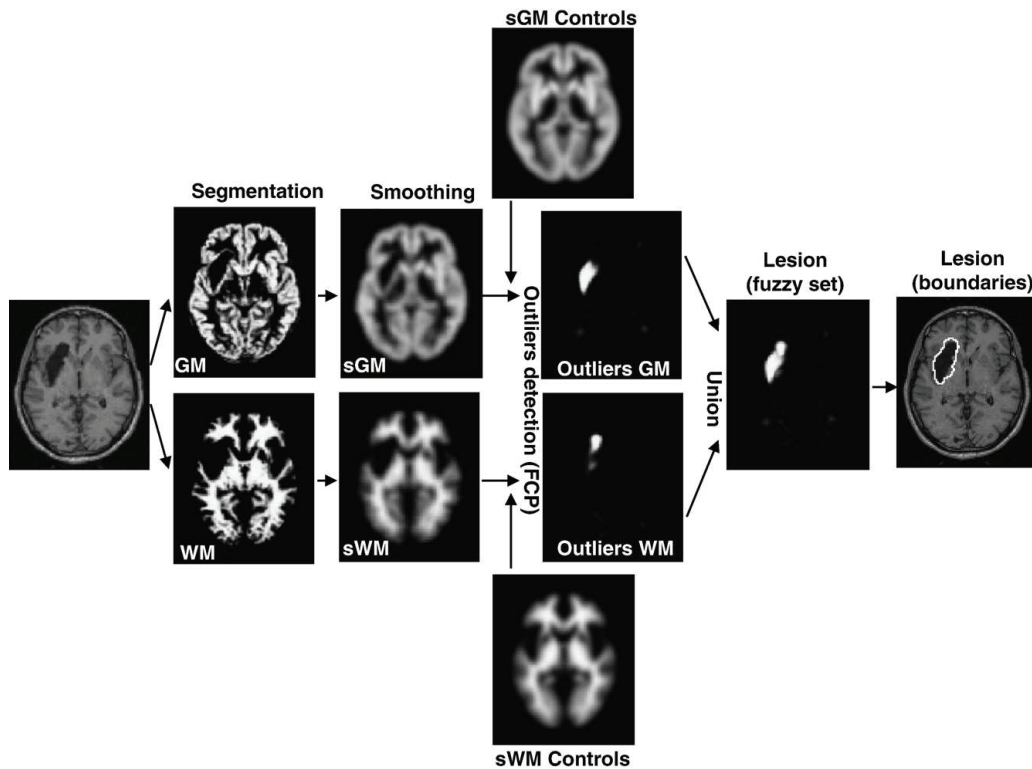


FIGURE III.4 – Vue d'ensemble de la méthode de détection de lésion proposée par SEGHIER et al., 2008. L'image est d'abord segmentée en différents groupes. La classification obtenue est lissée puis comparée à celle obtenue chez des patients sains. Les incohérences entre les deux classifications sont attribuées à la présence d'une lésion.

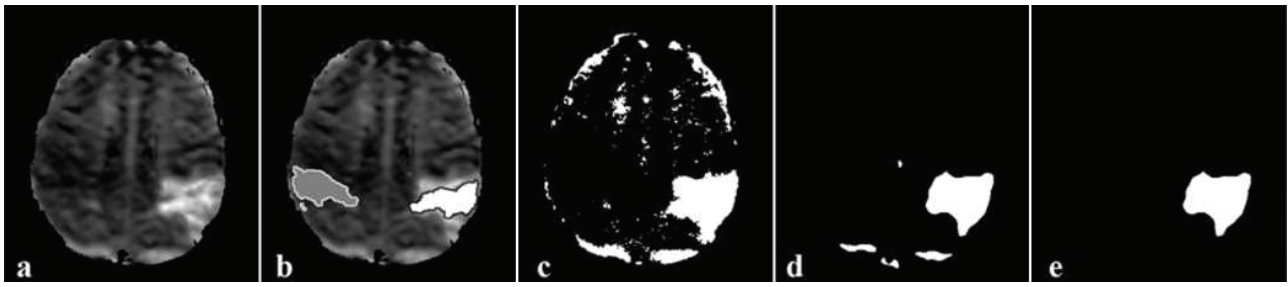


FIGURE III.5 – Application du modèle de Markov caché proposée par DWYER et al., 2008. (a) Cartographie du MTT. (b) Initialisation automatique de la lésion en repérant les valeurs anormales vis à vis de l'hémisphère controlatérale. (c) Seuillage. (d) Application du modèle de Markov caché. (e) Retraitement de la zone de lésion par critère de connectivité spatiale.

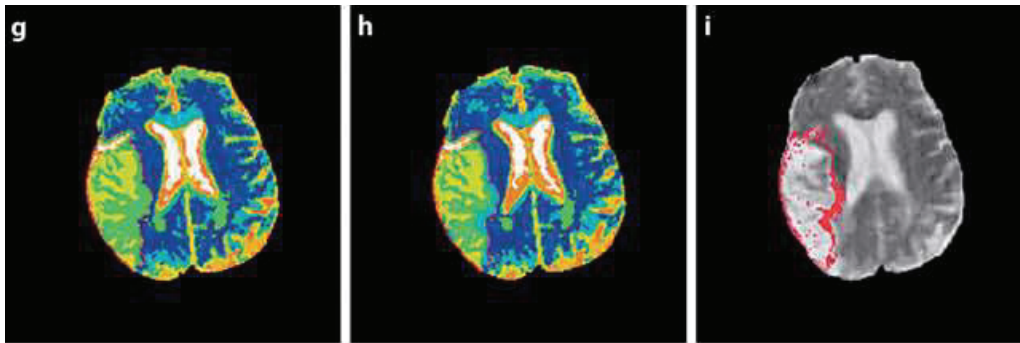


FIGURE III.6 – Classification obtenue avec l'algorithme ISODATA (SOLTANIAN-ZADEH et al., 2007). (g) Avec injection de Gadolinium. (h) Sans injection de Gadolinium. La figure (i) représente la lésion finale observée en T2 avec en rouge la différence entre les deux applications de l'algorithme ISODATA.

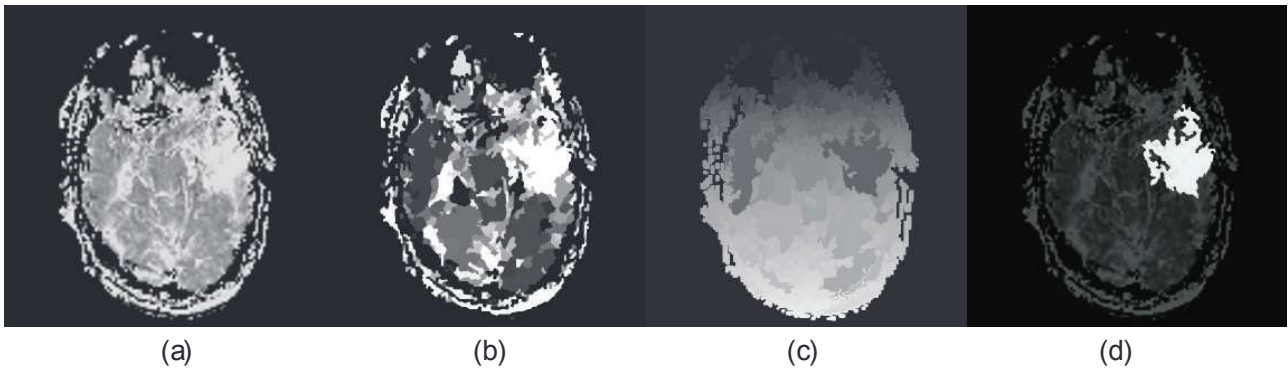


FIGURE III.7 – Application de l'algorithme 'Mean Shift' proposée par LI et al., 2009. (a) Cartographie initiale. (b) Groupes formés par l'algorithme 'Mean Shift'. (c) Groupes sélectionnés. (d) Mise en évidence du groupe sélectionné.



FIGURE III.8 – Exemple de retraitement de la zone de lésion proposée par MEILŪNAS et al., 2003. La zone de lésion (à gauche) est filtrée pour éliminer les groupes de voxels isolés (à droite).

Afin d'améliorer le rapport signal sur bruit des paramètres d'intensité, de nombreux travaux ont recouru à des méthodes de filtrage : filtrage gaussien (ARTZI et al., 2013; DASTIDAR et al., 2000; GILLEBERT et al., 2014), filtrage médian (LI et al., 2009), filtrage non linéaire (JACOBS et al., 2001b; LI et al., 2004; SOLTANIAN-ZADEH et al., 2007; FORBES et al., 2010b; FORBES et al., 2010a). Des méthodes corrigeant les gradients d'intensité causés par des hétérogénéités de champ magnétique lors de l'acquisition sont également employées (SOLTANIAN-ZADEH et al., 2007; FORBES et al., 2010a; FORBES et al., 2010b; MITRA et al., 2014; MAIER et al., 2015). Par ailleurs, plusieurs travaux mentionnent une ou plusieurs étapes de retraitement de la zone segmentée, généralement pour améliorer sa cohérence spatiale (JAMES et al., 2006; SHEN et al., 2008b; SHEN et al., 2010; ARTZI et al., 2013; MAH et al., 2014; MAIER et al., 2015) ou pour sélectionner les régions segmentées (HEVIA-MONTIEL et al., 2007; LI et al., 2009). MEILŪNAS et al., 2003, proposent d'ailleurs est un algorithme de retraitement d'une zone de lésion déjà segmentée. Il inclut notamment une étape de filtrage pour éliminer les voxels isolés (figure III.8 page 103) et une étape de lissage de la forme de la frontière.

La majorité des méthodes proposées sont des approches complètement automatiques. Néanmoins certaines méthodes nécessitent l'intervention de l'utilisateur pour indiquer une ou plusieurs zones servant à initialiser l'algorithme (DASTIDAR et al., 2000; JAMES et al., 2006; CONTIN et al., 2010), indiquer les régions d'intérêts parmi celles segmentées (LI et al., 2009) ou bien corriger la segmentation obtenue (MARTEL et al., 1999; JAMES et al., 2006; ARTZI et al., 2013). Cette dernière étape est particulièrement pertinente pour les méthodes déjà intégrées au sein d'un logiciel (DASTIDAR et al., 2000; CONTIN et al., 2010; ARTZI et al., 2013; HAAN et al., 2015; HANSEN et al., 2015).

Outils d'évaluation des approches de segmentation

Dans tous les articles retenus, la ou les méthodes de segmentation étaient comparées à la segmentation manuelle. Certaines articles proposent une comparaison de plusieurs approches de segmentation automatique (USINSKAS et al., 2003; GHOSH et al., 2014) ou bien de comparer des approches automatiques et semi-automatiques (WILKE et al., 2011) par rapport à la segmentation manuelle. Il n'était mention dans aucun article de méthode de segmentation automatique de référence.

En raison du faible nombre d'études comparatives, il pourrait être intéressant de comparer les approches d'après les performances reportées dans les articles. Toutefois, pour que la comparaison soit fiable, il faudrait évaluer les approches de segmentation sur un même jeu de données. En effet, les performances d'une méthode de segmentation peuvent varier considérablement suivant les patients considérés, certains ayant des lésions plus difficiles à segmenter que d'autres, suivant le contraste de l'image, la présence d'artéfacts ou la forme de la lésion. De plus toutes les approches n'ont pas été conçues et évaluées sur le même type de lésion. Dans la majeure partie des cas, il s'agissait d'identifier la lésion en phase aigüe ou

chronique. Toutefois, certains travaux se sont intéressés à la lésion en perfusion (JACOBS et al., 2001b; MOURIDSEN et al., 2004; JAMES et al., 2006; SOLTANIAN-ZADEH et al., 2007; DWYER et al., 2008; CONTIN et al., 2010). À noter que quelques articles évaluent leur approche sur des données disponibles en ligne (BRAUN et al., 2002; SHEN et al., 2010; FORBES et al., 2010b) ce qui rend possible une comparaison.

De nombreux critères existent dans la littérature pour comparer deux segmentations : voir TAHA et al., 2015, pour une revue critique des différents critères. Certains ont déjà été présentés dans la partie 2.2 page 20. À noter que les critères d'aire sous la courbe sont ici peu pertinents puisque la segmentation doit permettre de décider d'une partition binaire. Même si de nombreux algorithmes proposent des classifications floues, ces classifications floues sont généralement retraitées pour aboutir à une classification binaire avant évaluation. En suivant le travail de TAHA et al., 2015, nous regroupons les principaux indices relevés dans la littérature en 3 catégories :

- **les indices évaluant le recouvrement volumique.** Ils examinent la correspondance voxel à voxel entre deux segmentations. La plupart des articles reportent au moins un indice de ce type, avec principalement : l'indice de Dice (GILLEBERT et al., 2014; MAIER et al., 2015; MITRA et al., 2014; MOURIDSEN et al., 2013; PRAKASH et al., 2006; WILKE et al., 2011; FORBES et al., 2010a; FORBES et al., 2010b; SEGHER et al., 2008; HAAN et al., 2015; HANSEN et al., 2015), l'indice de similarité (LI et al., 2004; HEVIA-MONTIEL et al., 2007; SHEN et al., 2008b; SHEN et al., 2010; SAAD et al., 2012; MAH et al., 2014; GHOSH et al., 2014), l'indice de Jaccard (HAAN et al., 2015). L'utilisation de la sensibilité et de la spécificité est aussi très fréquente (BRAUN et al., 2002; PRAKASH et al., 2006; SHEN et al., 2008b; SEGHER et al., 2008; SHEN et al., 2010; SAAD et al., 2012; GHOSH et al., 2014; GILLEBERT et al., 2014; MAH et al., 2014; MITRA et al., 2014; HANSEN et al., 2015). En revanche, seuls quelques articles reportent la précision et le rappel (CHAWLA et al., 2009; MAIER et al., 2015).
- **les indices évaluant la précision volumique.** Ils comparent les volumes obtenus par deux segmentations indépendamment de la correspondance spatiale entre les segmentations. Les volumes peuvent être comparés directement (STEIN et al., 2001; JAMES et al., 2006; HEVIA-MONTIEL et al., 2007; DALCA et al., 2014), en valeur relative (ARAKAWA et al., 2006; DWYER et al., 2008; WEINMAN et al., 2003; MITRA et al., 2014; ARTZI et al., 2013; GHOSH et al., 2014), en évaluant leur corrélation (DALCA et al., 2014; JACOBS et al., 2001b; JAMES et al., 2006; MITRA et al., 2014; SOLTANIAN-ZADEH et al., 2007) ou en utilisant un graphique de Bland et Altman (MOURIDSEN et al., 2013; HANSEN et al., 2015). Quelques articles évaluent aussi la corrélation entre le volume segmenté et un score clinique (DASTIDAR et al., 2000; JACOBS et al., 2001b).
- **les indices évaluant la précision de la délimitation.** Ils comparent la position des frontières des deux segmentations. La distance de Hausdorff est utilisée (MAIER et al.,

2015) ainsi que la distance symétrique moyenne ("average symmetric surface distance", ARTZI et al., 2013 ; MAIER et al., 2015 ; MITRA et al., 2014).

La plupart des études évaluent la segmentation à l'aide de plusieurs indices. En effet, les différents indices reflètent plusieurs façons de comparer deux segmentations, chacune ayant des avantages et des inconvénients (voir tableau 4 de TAHA et al., 2015). En imagerie médicale, l'application clinique permet de déterminer l'angle de comparaison en fonction duquel il convient de choisir la ou les indices de performance. Dans le cadre de l'AVC, une bonne estimation du volume est primordial ainsi qu'un positionnement correcte de la lésion. En revanche des imprécisions sur la position de la frontière sont acceptables. Il semble donc utile de reporter à la fois un indice de recouvrement volumique et un indice de précision volumique. A noter qu'au sein de chaque type d'indice, les indices souvent très corrélés entre eux (voir figure 3 de l'article de TAHA et al., 2015). Il n'est donc pas forcément pertinent de reporter plusieurs indice de chaque type. Il y a toutefois une différence majeure entre les indices prenant en compte les vrais négatifs (comme la spécificité) et les indices les ignorants (comme l'indice de Dice). Comme nous l'avons vu dans la section 2.4 page 26, le deuxième type d'indice va mieux refléter les performances de la méthode de segmentation lorsque la prévalence est faible, ce qui est généralement le cas pour l'AVC.

1.3 Objectif

L'objectif du travail est de disposer d'un algorithme de segmentation de lésion pour l'AVC. Deux séquences sont classiquement utilisées : le DWI à l'admission du patient et le T2 FLAIR à l'IRM finale. Comme indiqué dans la littérature et illustré à la figure III.1, ces séquences sont les plus informatives pour identifier la lésion. En phase aigüe, la séquence en perfusion apporte une information complémentaire à celle en diffusion mais son acquisition n'est pas systématique. En phase chronique, la séquence en diffusion a une moins bonne sensibilité à la lésion que le T2 FLAIR. N'étant pas affectée par la maladie de la substance blanche, elle garde néanmoins un intérêt en raison de sa bonne spécificité. L'algorithme de segmentation doit donc permettre une caractérisation multivariée de la lésion tout en étant robuste aux éventuels artéfacts que pourrait causer la maladie de la substance blanche. Il sera évalué sur sa capacité à segmenter la lésion sur les cartographies T2 FLAIR : d'une part puisqu'elle est requise pour entrainer des modèles prédictifs et d'autre part car le T2 FLAIR est plus difficile à segmenter que le DWI en raison de la présence de nombreux artéfacts.

Choix du type d'approche

Lors de la réalisation de ce travail, les articles de DALCA et al., 2014 ; MITRA et al., 2014 ; MAIER et al., 2015, n'étaient pas encore publiés. Il n'y avait donc pas d'approche dans la littérature robuste aux artéfacts causés par la maladie de la substance blanche. Or ces artéfacts rendent la segmentation peu spécifique : des corrections importantes sont alors nécessaires.

Cette maladie étant courante chez les patients âgés, une nouvelle méthode de segmentation apparaissait nécessaire. Étant donné la grande hétérogénéité de forme et de contraste entre les lésions, nous souhaitons une approche qui puisse s'adapter aux spécificités de chaque patient.

Nous avons écarté les approches d'analyse d'intensité qui nous semblaient reposer sur des choix arbitraires et peu robustes. Nous avons également écarté les méthodes d'apprentissage supervisé car elles impliquent une opération de normalisation spatiale. Cette étape peut être source d'erreurs (voir par exemple THIRION et al., 2006) et n'est pas requise par les autres méthodes. Parmi les approches restantes, les modèles de mélange nous ont paru les plus prometteurs. Ils offrent en effet un cadre de modélisation très flexible tout en permettant d'estimer l'ensemble des paramètres du modèle.

1.4 Le modèle de mélange pour la segmentation d'image

Dans cette partie nous présentons une brève introduction au modèle de mélange et à ses extensions pour la segmentation d'image. Cette introduction reprend certains éléments des chapitres 1 et 2 de BLANCHET, 2007, qui fournit une description plus complète des modèles de mélange sur champ de Markov.

Notations

Nous considérerons un seul patient pour lequel on dispose de n sites d'observation. L'ensemble des sites sera noté \mathcal{I} et $\mathcal{I} \setminus i$ désignera l'ensemble des sites à l'exception du site i . À chaque site est associée une variable aléatoire Y_i (potentiellement multidimensionnelle) dont on observe une réalisation. Nous supposerons que les sites appartiennent à un nombre fini de structures cérébrales aussi appelées groupes. Ce nombre sera noté G . L'appartenance des sites aux différents groupes sera notée à travers un vecteur de variables aléatoires indicatrices : $\boldsymbol{\xi} = (\xi_i, \dots, \xi_i)$. La g -ième composante de ξ_i , notée $\xi_{i,g}$, vaut 1 lorsque le site i appartient au groupe g et 0 sinon. Chaque site appartient à un unique groupe : $\forall i \in \{1, \dots, n\} \sum_{g=1}^G \xi_{i,g} = 1$. Par convention $\xi_i = g$ indiquera que la g -ième composante du vecteur ξ_i vaut 1 et que toutes les autres sont nulles (le site i appartient au groupe g). Enfin, comme pour $\boldsymbol{\xi}$ nous noterons $\mathbf{Y} = (Y_1, \dots, Y_n)$.

REMARQUE : Dans notre application les sites d'observation correspondent aux voxels. Nous utiliserons une intensité \mathbf{Y} bivariée composée du T2 FLAIR et du T2 FLAIR normalisé par la valeur controlatérale. Ce dernier paramètre permet d'exploiter l'asymétrie qu'engendre la lésion dans la distribution spatiale d'intensité et l'approximative symétrie de la maladie de la substance blanche par rapport au plan sagittal médian. Enfin, nous considérerons 4 groupes : la lésion, la substance blanche, la substance grise et le liquide cérébro-spinal.

Le modèle de mélange

Le modèle de mélange modélise l'intensité observée par un mélange de G distributions, chacune ayant des paramètres spécifiques notés θ_g . En notant Θ l'ensemble des paramètres du modèle, la distribution de probabilité des observations s'exprime :

$$\mathbb{P}[\mathbf{Y}|\Theta] = \sum_{\xi \in \Omega} \mathbb{P}[\mathbf{Y}|\xi, \Theta] \mathbb{P}[\xi|\Theta]$$

où Ω indique l'ensemble des configurations possibles pour l'appartenance aux groupes. Nous ferons l'hypothèse de bruit indépendant :

$$\begin{aligned} \text{(i)} \quad & \mathbb{P}[\mathbf{Y}|\xi] = \prod_{i=1}^n \mathbb{P}[Y_i|\xi] \\ \text{(ii)} \quad & \forall i \in \{1, \dots, n\} \quad \mathbb{P}[Y_i|\xi] = \mathbb{P}[Y_i|\xi_i] \end{aligned}$$

Ainsi nous supposons que (i) conditionnellement à l'appartenance aux groupes, les variables aléatoires associées aux intensités de chaque groupe sont indépendantes (ii) Connaissant le groupe d'un site, l'intensité observée est indépendante du groupe des sites voisins.

REMARQUE : La première hypothèse peut être mis en défaut par exemple en présence d'hétérogénéité du champ magnétique lors de l'acquisition : l'hétérogénéité va générer un gradient d'intensité et donc une corrélation entre les intensités voisines, indépendamment de l'appartenance aux groupes. La seconde hypothèse implique par exemple l'absence de contamination de l'intensité d'un voxel par le tissu voisin (effet de volume partiel).

Sous cette hypothèse de bruit indépendant, la distribution de probabilité peut se ré-écrire :

$$\mathbb{P}[\mathbf{Y}|\Theta] = \sum_{\xi \in \Omega} \left(\prod_{i=1}^n \mathbb{P}[Y_i|\xi_i, \theta_{\xi_i}] \right) \mathbb{P}[\xi|\Theta] \quad (\text{III.1})$$

où θ_{ξ_i} désigne θ_g lorsque le site i appartient au groupe g (i.e. ξ_i est non nul à la g -ième position). Le modèle de mélange peut être décomposé en deux termes : un premier terme $\prod_{i=1}^n \mathbb{P}[Y_i|\xi_i, \theta_{\xi_i}]$ correspondant au modèle d'intensité. À chaque groupe g est associé une distribution de paramètres θ_g . Ainsi, ce terme correspond à la densité de la distribution évaluée en Y_i , notée $f_{\theta_g}(Y_i)$. Les distributions sont souvent choisies gaussienne : θ_g correspond alors à l'espérance et la variance de la loi gaussienne. Par exemple, dans le cas univarié :

$$\mathbb{P}[Y_i|\xi_i, \theta_{\xi_i}] = f_{\theta_{\xi_i}}(Y_i) = \frac{1}{\sqrt{2\pi}\sigma_{\xi_i}} \exp\left(-\frac{(Y_i - \mu_{\xi_i})^2}{2\sigma_{\xi_i}^2}\right)$$

Le second terme correspond à l'a priori sur la distribution des groupes donné par le modèle spatial. Dans le cadre de la segmentation d'image, toutes les distributions de groupes ne sont pas équiprobables. On s'attend généralement à ce que les observations du même groupe s'agglomèrent.

Ce terme permet de spécifier la géométrie attendue pour les groupes, qui peut varier d'une application à l'autre (tout comme le choix de la distribution d'intensité). Classiquement, le modèle de Potts est utilisé. Il s'agit d'un modèle simple permettant de corriger un bruit ponctuel dans la mesure de l'intensité.

Notion de voisinage

Afin de mettre en place un modèle spatial, il est nécessaire de définir les relations spatiales qui existent entre les sites. Pour cela, nous introduisons la notion de système de voisinage. Ce système définit les relations de voisinage qui existent entre chaque site et tous les autres sites. La relation de voisinage est supposée symétrique, c'est à dire que si un site i est voisin d'un site j , la réciproque est vraie. Par ailleurs, nous définirons plusieurs ordres de voisinage, chaque ordre correspondant à plusieurs degrés de proximité dans la relation de voisinage. Les plus proches voisins peuvent par exemple former le premier ordre de voisinage : pour un réseau de sites disposés régulièrement sur une grille de dimension deux, chaque site a alors 4 voisins de premier ordre. Le second ordre peut être défini comme les plus proches voisins suivants et ainsi de suite (voir figure III.9). D'autres définition du voisinage sont possibles, comme par exemple considérer pour chaque ordre, l'ensemble de sites dont la distance est inférieure à une valeur donnée, et qui ne sont pas déjà inclus dans un ordre inférieur.

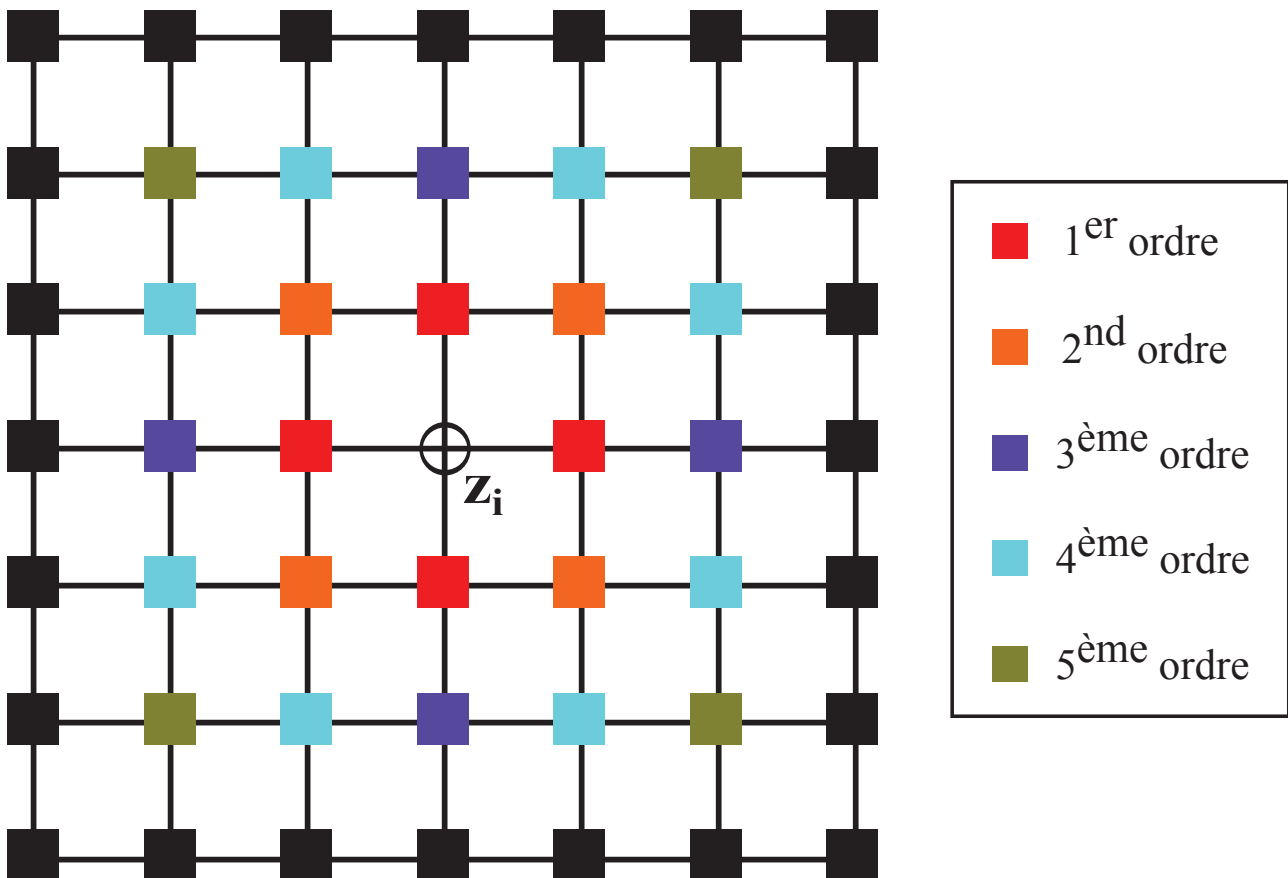


FIGURE III.9 – Exemple de système de voisinage en deux dimensions.

Par la suite, nous désignerons par $\mathbf{N}(i)$ l'ensemble des voisins du site i et par $\mathbf{N}_c(i)$ sa restriction aux voisins d'ordre c . En théorie rien ne limite la portée du voisinage. Toutefois, il est peu réaliste de considérer que les sites lointains aient une influence (typiquement des sites situés dans des hémisphères cérébraux différents ont peu d'interactions). Pour cela on limite le voisinage de chaque site à un ordre maximum noté C_{max} . Cette hypothèse nous permettra d'assimiler l'ensemble des sites à un champ de Markov et d'en préciser la forme de la distribution de probabilité.

Distribution de probabilité sur un champ de Markov

En reprenant les notations précédemment introduites, ξ est un champ de markov associé au graphe de voisinage \mathbf{N} si :

- (i) $\forall i \in \{1, \dots, n\}$ et $\forall \mathbf{x} \in \Omega$ $\mathbb{P}[\xi_i = x_i | \xi_{\mathcal{I} \setminus \{i\}} = \mathbf{x}_{\mathcal{I} \setminus \{i\}}] = \mathbb{P}[\xi_i = x_i | \xi_{\mathbf{N}(i)} = \mathbf{x}_{\mathbf{N}(i)}]$
- (ii) $\forall \mathbf{x} \in \Omega$ $\mathbb{P}[\xi = \mathbf{x}] > 0$

(i) indique que chaque variable aléatoire dépend uniquement d'un nombre limité de sites, ce qui correspond à l'hypothèse faite dans la section précédente où l'on limite l'ordre maximal de voisinage. (ii) impose que chaque configuration ait une probabilité. Elle est bien vérifiée sur notre application puisque, a priori, chacun des sites peut appartenir à n'importe quel des groupes.

Le théorème de Hammersely-Clifford (BESAG, 1974) va permettre de spécifier la distribution de probabilité jointe. En effet, ce théorème indique que sur un champ de Markov, la distribution de probabilité jointe est une distribution de Gibbs :

$$\mathbb{P}[\xi] = \frac{1}{Z} \exp \left(- \sum_{c \in \mathcal{C}} V_c(\xi_c) \right)$$

où Z est une constant de normalisation, V_c est la fonction potentiel associée aux cliques d'ordre c et \mathcal{C} désigne l'ensemble des cliques. Une clique est un ensemble de points tous voisins entre eux. Par exemple, les cliques d'ordre 1 sont des singletons, les cliques d'ordre deux des paires de sites voisins entre eux, et les cliques d'ordre trois les triplets de sites tous voisins entre eux.

La distribution obtenue a une forme très générale. Le modèle de Potts est un cas particulier de cette distribution où l'on va préciser la forme des potentiels.

Le modèle de Potts

Le modèle de Potts repose principalement sur deux hypothèses :

- (i) seuls les potentiels associés aux cliques d'ordre un ou deux sont non nuls.
- (ii) la forme du potentiel est identique en tout site (hypothèse de stationnarité).

En pratique il existe une troisième hypothèse de voisinage local. En effet, la portée du voisinage est généralement restreinte aux premiers ordres de voisinage (ordre 1 ou 2). Par exemple, dans l'article de DWYER et al., 2008, les 26 plus proches voisins sont utilisés. De même, l'algorithme implémenté par FENG et al., 2011, permet d'utiliser un voisinage local (en trois dimensions) composé soit de 6 voisins, de 18 voisins ou de 26 voisins. Par la suite, nous désignerons par $N_{local}(i)$ le voisinage local du site i .

REMARQUE : Dans notre application ce voisinage local contiendra 10 voisins : les 8 sites les plus proches dans le plan et le voisin le plus proche dans le plan inférieur et supérieur. Ce choix est justifié par une résolution axiale de l'IRM inférieure à la résolution dans le plan.

Sous ces hypothèses, nous pouvons ré-écrire la distribution de probabilité :

$$\mathbb{P}[\boldsymbol{\xi}] = \frac{1}{Z} \exp \left(- \sum_{i=1}^n V_1(\xi_i) - \sum_{i=1}^n \sum_{j \in N_{local}(i)} V_2(\xi_i, \xi_j) \right)$$

Le premier potentiel V_1 peut permettre de gérer un a priori sur l'appartenance à chaque groupe. Il est aussi appelé champ externe en référence à la mécanique statistique où il permettrait de prendre en compte la présence d'un champ magnétique externe rendant plus probable certains états de spin. Le second potentiel va permettre de préciser les configurations de voisinage les plus probables. Pour chacune des configurations, un coefficient sera attribué. La configuration où un site appartient au groupe g et son voisin au groupe g' aura un coefficient noté $\rho_{2,g,g'}$. Ainsi en introduisant la matrice $\boldsymbol{\rho}_2 = (\rho_{2,g,g'})_{(g,g') \in \{1, \dots, G\}^2}$, le second potentiel peut se ré-écrire :

$$-V_2(\xi_i, \xi_j) = {}^t \xi_i \boldsymbol{\rho}_2 \xi_j$$

Le modèle de Potts usuel suppose que $\boldsymbol{\rho}_2$ est proportionnel à la matrice identité : $\boldsymbol{\rho}_2 = \rho_2 \mathbf{I}_{G,G}$ avec $\rho_2 > 0$. Cela permet de favoriser les voisinages homogènes de façon égale entre les groupes et de défavoriser les voisinages hétérogènes. Se reporter à la figure E de l'article page 152 pour voir l'influence du paramètre ρ_2 sur le résultat d'une simulation du modèle de Potts. En renommant $-V_1$ par $V_{externe}$, $-V_2$ par V_{local} et ρ_2 par ρ_{local} nous pouvons ré-écrire la distribution associée au modèle de Potts :

$$\mathbb{P}[\boldsymbol{\xi}] = \frac{1}{Z} \exp \left(\sum_{i=1}^n V_{externe}(\xi_i) + \rho_{local} \sum_{i=1}^n \sum_{j \in N_{local}(i)} {}^t \xi_i \xi_j \right)$$

Le modèle de mélange spatial

Les modèles de mélange pour la segmentation d'image utilisent classiquement le modèle de Potts pour spécifier l'a priori spatial (voir par exemple l'article de KABIR et al., 2007). La vraisemblance peut alors se ré-écrire :

$$\mathbb{P}[\mathbf{Y}|\boldsymbol{\Theta}] = \sum_{\boldsymbol{\xi} \in \Omega} \left(\prod_{i=1}^n f_{\theta_{\xi_i}}(Y_i) \right) \frac{1}{Z} \exp \left(\sum_{i=1}^n V_{\text{externe}}(\xi_i) + \rho_{\text{local}} \sum_{i=1}^n \sum_{j \in \mathbf{N}_{\text{local}}(i)} t_{\xi_i \xi_j} \right)$$

où le champ externe est souvent choisi nul (tous les groupes sont équiprobables) ou alors fixé au logarithme de la prévalence du groupe (la prévalence est utilisée comme a priori à l'appartenance aux groupes).

Limites du modèle spatial et améliorations existantes

Le modèle de Potts n'offre pas une modélisation satisfaisante dans le cadre de la segmentation d'image. Cependant, même si ce constat est ancien (MORRIS et al., 1996; TJELMELAND et al., 1998), ce modèle est encore largement utilisé en segmentation en raison de son efficacité et de sa simplicité. Les deux principaux reproches fait au modèle de Potts sont :

- **il ne préserve pas le détail des images.** En effet le modèle de Potts pénalise les voisinages hétérogènes. Ceux-ci peuvent être causé par du bruit dans la mesure, qu'il convient de corriger, ou bien par l'existence d'une frontière, qu'il convient de préserver. Le modèle de Potts va favoriser les frontières courtes et tend donc à dégrader niveau de détail des images. Comme l'indiquent DESCOMBES et al., 1996, "ce modèle ne fait pas la distinction entre contour et bruit".
- **il ne permet pas de former des structures cohérentes à grande échelle.** Ce défaut est particulièrement problématique en reconnaissance de forme où l'objet peut-être partiellement masqué ou avoir une intensité irrégulière. Comme discuté dans la section 1.1 page 94, en imagerie cérébrale c'est la présence d'artéfacts et l'hétérogénéité du signal qui va nuire à la cohésion à grande échelle.

De approches ont donc été proposées pour palier à ces défauts. Nous en mentionnerons trois qui vont compléter ou généraliser le modèle de Potts :

- Afin de distinguer le bruit des frontières, GEMAN et al., 1984, proposent de modéliser explicitement les frontières. Le principe étant d'introduire une nouvelle variable aléatoire L indiquant pour chaque point s'il s'agit d'un point frontière. L'homogénéité des cliques relatives aux points frontières n'est alors pas évaluée, ce qui permet de préserver les détails.

- Afin de préserver les détails, DESCOMBES et al., 1996, généralisent le modèle de Potts en utilisant des cliques d'ordre supérieur à 2. Ce nouveau modèle, appelé "chien-modèle" considère les cliques formant une matrice 3 lignes et 3 colonnes (voir figure III.10 à gauche). Cela va permettre de différencier les configurations spatiales associées au bruit de celles associées aux contours ou aux lignes. Typiquement une clique où l'on observe des groupes isolés se voit attribuer un potentiel faible tandis qu'une clique où l'on observe une structure de ligne (voir figure III.10 à droite) aura un potentiel élevé, même si la clique n'est pas homogène.
- Pour favoriser la consistance spatiale à grande échelle, des modèles de Markov considérant des cliques d'ordre important ont été proposés (aussi appelés "high-order markov model"). Parmi les travaux existants nous avons retenu celui de KOHLI et al., 2009, qui proposent une extension du modèle de Potts en ajoutant un potentiel de cohérence régional. La distribution de Gibbs contient alors trois potentiels : le champ externe, le potentiel d'interaction entre les cliques d'ordre 2 et le potentiel de cohérence régional. Ce dernier évalue l'homogénéité des groupes sur des "superpixels" définis au préalable :

$$-V_{régional}(\xi_{superpixel}) = \begin{cases} \bar{N}_i(\xi_{superpixel}) \frac{\gamma_{max}}{Q} & \text{si } \bar{N}_i(\xi_{superpixel}) \leq Q \\ \gamma_{max} & \text{sinon} \end{cases}$$

où $\xi_{superpixel}$ désigne les indicatrices de groupe des sites appartenant au supervoxel, $\bar{N}_i(\xi_{superpixel})$ est le nombre de sites dont le groupe est différent du groupe majoritaire dans le "superpixel", Q et γ_{max} sont des paramètres. Le signe "-" devant le potentiel régional est nécessaire puisque, contrairement à KOHLI et al., 2009, nous avons pris la convention que les potentiels élevés favorisent la configuration spatiale (en désignant par $V_{régional}$ l'opposé du potentiel de la clique). Avec ce potentiel régional, les "superpixels" homogènes sont favorisés ce qui permet de donner une cohérence à grande échelle à la segmentation. À noter que le potentiel régional correspond ici à un potentiel d'ordre le nombre de sites présents dans le "superpixel". Le forme du potentiel est donc variable suivant le site considéré.

Afin de définir les "superpixels", KOHLI et al., 2009, proposent d'utiliser l'algorithme "Mean Shift" avec pour noyau le produit d'un noyau spatial et d'un noyau d'intensité. Les deux noyaux ont un paramètre de bande passante propre et utilisent une distance euclidienne pour agglomérer les variables spatiales et les variables d'intensité. L'algorithme "Mean Shift" est un algorithme permettant d'identifier les maxima de densité puis d'en déduire une partition des observations en rattachant chaque observation à un maximum (voir figure III.11). L'utilisation d'un noyau spatial permet d'imposer une cohérence spatiale à la partition.

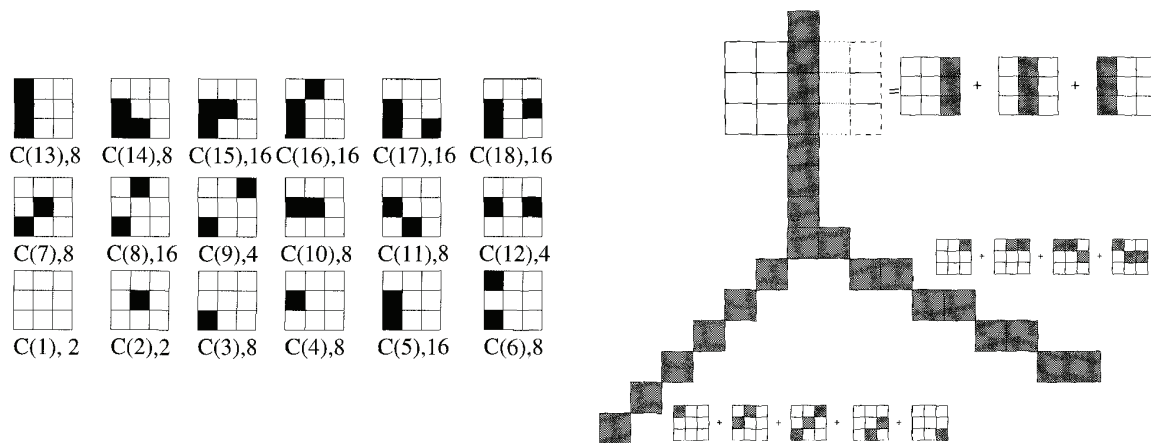


FIGURE III.10 – Illustration du chien modèle. À gauche : 13 premières cliques considérées dans le chien modèle. À droite : définition des lignes dans le chien modèle. Figures extraites de DESCOMBES et al., 1996.

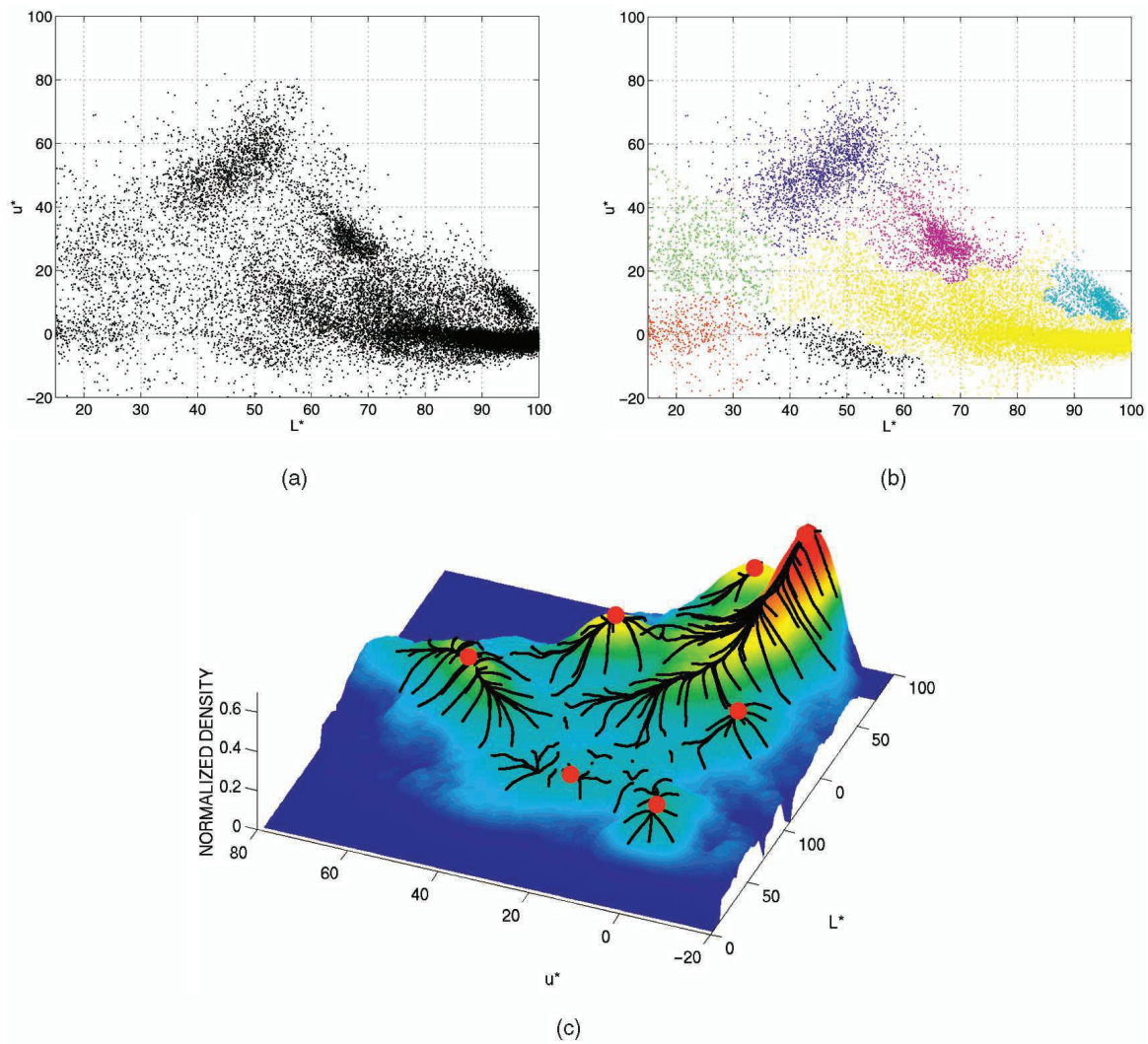


FIGURE III.11 – Exemple d'application de l'algorithme "Mean Shift" : (a) Nuage de points dans l'espace d'intensité. (b) Partitionnement des observations proposé par l'algorithme "Mean Shift". (c) Maxima d'intensité trouvés par l'algorithme. Figure extraite de COMANICIU et al., 2002

Même si ces approches ont montré leur efficacité sur les applications proposées, aucune n'a été conçue pour gérer des artéfacts d'intensité. Afin de corriger ces artéfacts, nous allons proposer une nouvelle extension du modèle de Potts définissant un potentiel additionnel au potentiel local.

1.5 Extension du modèle de Potts proposée

Nous introduisons un nouveau potentiel dans le modèle de Potts afin de pénaliser les formes de lésion allongées ou tortueuses et de favoriser les formes sphériques. Comme KOHLI et al., 2009, ce potentiel intégrera de l'information régionale. Toutefois il n'utilisera que des cliques d'ordre 2. Nous allons donc considérer deux ensembles de voisinage pour chaque site i : un voisinage local noté comme précédemment $\mathbf{N}_{local}(i)$ et un voisinage régional noté $\mathbf{N}_{régional}(i)$. Nous supposerons que le voisinage régional est composé de sites dont l'ordre de voisinage varie entre 1 et C_{max} avec au moins un site pour chaque ordre.

Nous allons tout d'abord normaliser la contribution des voisins. Pour le voisinage local, nous considérerons un poids $w_{i,j}^{local}$ pour le voisin j du site i . Chaque poids sera fixé égal à l'inverse du nombre de voisins contenu dans le voisinage local du site¹ :

$$w_{i,j}^{local} = \frac{1}{\sum_{j \in \mathbf{N}_{local}(i)} 1}$$

Pour le voisinage régional, nous considérerons un poids $w_{i,j}^c$ pour le voisin j d'ordre c du site i . Chaque poids sera fixé égal à l'inverse du nombre de voisins contenu dans le voisinage d'ordre c du site i :

$$w_{i,j}^c = \frac{1}{\sum_{j \in \mathbf{N}_c(i)} 1}$$

Définissons $N_{local}(\xi_i, \boldsymbol{\xi}_{\mathbf{N}_{local}(i)})$ le pourcentage de cliques homogènes présentes dans le voisinage local du site i :

$$N_{local}(\xi_i, \boldsymbol{\xi}_{\mathbf{N}_{local}(i)}) = \sum_{j \in \mathbf{N}_{local}(i)} w_{i,j}^{local} {}^t \xi_i \xi_j$$

Nous redéfinissons le potentiel local comme :

$$V_{local}(\boldsymbol{\xi}) = \sum_{i=1}^n N_{local}(\xi_i, \boldsymbol{\xi}_{\mathbf{N}_{local}(i)})$$

Définissons à présent $N_c(\xi_i, \boldsymbol{\xi}_{\mathbf{N}_c(i)})$ le pourcentage de cliques homogènes relatives au site i à l'ordre de voisinage c :

$$N_c(\xi_i, \boldsymbol{\xi}_{\mathbf{N}_c(i)}) = \sum_{j \in \mathbf{N}_c(i)} w_{i,j}^c {}^t \xi_i \xi_j$$

1. D'autres choix sont possibles pour normaliser à 1 la contribution de l'ensemble du voisinage. Par exemple prendre pour chaque poids l'inverse de l'ordre de voisinage du site voisin, multiplié par un coefficient. Ce coefficient est déterminé en imposant une somme des poids égale à 1 sur le voisinage.

Le potentiel régional est défini comme la moyenne sur les différents ordre de voisinage du pourcentage de cliques homogènes :

$$V_{régional}(\xi) = \sum_{i=1}^n \frac{1}{C_{max}} \sum_{c=1}^{C_{max}} N_c(\xi_i, \xi_{N_c(i)})$$

Il peut donc être vu comme une moyenne du potentiel local à différentes échelles. La figure III.12 illustre le calcul de chacun de ces deux potentiels pour un site. Considérons une partition de l'espace en deux groupes, où l'un des groupes est disposé selon une ligne (en rouge sur la figure nommée "Forme"). Considérons un site situé sur cette ligne (représenté par un rectangle vert sur la figure). Son potentiel local se calcul comme le pourcentage de voisins d'ordre 1 qui appartient à son groupe : il y en a 5 sur 8 d'où une valeur de 0,625 et donc une intensité moyenne du potentiel local (représentée en jaune sur la figure). Pour calculer le potentiel régional, il faut aussi évaluer l'homogénéité des cliques correspondant aux ordres de voisinage supérieur : à l'ordre 2 il y a 12 voisins dont 4 de son groupe d'où la valeur de 0,333. En moyennant les valeurs obtenues pour les 10 premiers ordre de voisinage on obtient une valeur proche de 0 (représentée en bleu clair). Comme on peut le voir ici, le potentiel régional prend des valeurs faibles pour les artéfacts linéaires. Or si le site était associé au groupe bleu, le potentiel régional prendrait des valeurs élevées. Ainsi, le potentiel régional permet de mettre un a priori plus favorable au groupe bleu qu'au groupe rouge et donc de corriger l'artéfact.

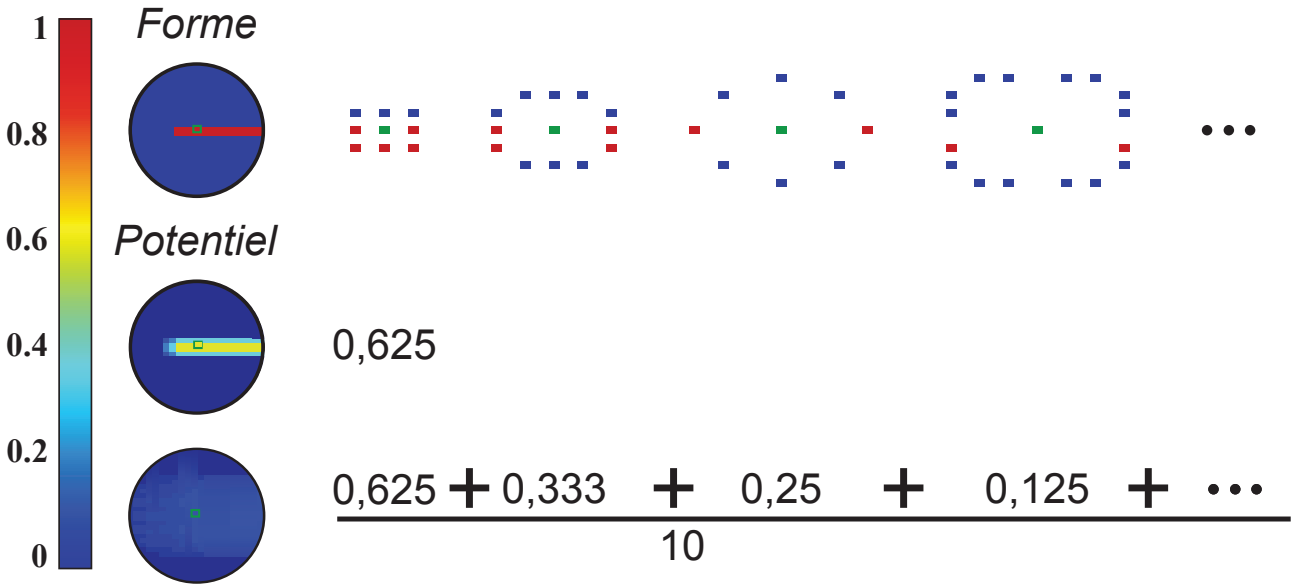


FIGURE III.12 – Méthodes de calcul du potentiel local (seconde ligne) et régional (troisième ligne) sur une forme (première ligne). Chaque colonne correspond à un ordre de voisinage.

En considérant un champ externe nul, le modèle de Potts se ré-écrit alors :

$$\mathbb{P}[\xi] = \frac{1}{Z} \exp(\rho_{local} V_{local}(\xi) + \rho_{régional} V_{régional}(\xi))$$

Il reste à définir C_{max} , l'ordre de voisinage maximal. Ce paramètre fixe l'échelle de régularisation de la forme de la lésion. Pour pouvoir corriger un artéfact linéaire, il faut que C_{max} soit plus grand que la largeur de la ligne (en pratique un facteur au moins égal à 2 est souhaitable). Toutefois, un C_{max} trop grand peut causer une correction excessive comme par exemple gommer des protubérances correspondant réellement à la lésion. À l'extrême, un C_{max} excédant la taille de la lésion va défavoriser l'appartenance à la lésion des voxels de la frontière et donc faire échouer la segmentation. Idéalement C_{max} devrait être fixé par l'utilisateur en fonction de son appréciation de l'échelle adaptée. Toutefois cela nuit au caractère automatique de la méthode, d'autant que ce choix n'est pas forcément évident. Aussi nous proposons d'estimer la valeur de C_{max} en fonction de l'étendue du groupe considéré : il est calculé comme la distance moyenne au barycentre. Si le groupe est spatialement fragmenté, il est aussi possible d'estimer et d'utiliser une valeur de C_{max} par groupe spatial (voir annexe F de l'article page 155).

1.6 Estimation du modèle de segmentation

Nous pouvons écrire la distribution de probabilité associée au modèle de mélange proposé :

$$\begin{aligned}\mathbb{P}[\mathbf{Y}|\boldsymbol{\Theta}] &= \sum_{\boldsymbol{\xi} \in \Omega} \left(\prod_{i=1}^n f_{\theta_{\xi_i}}(Y_i) \right) \frac{1}{Z} \exp(\rho_{local} V_{local}(\boldsymbol{\xi}) + \rho_{régional} V_{régional}(\boldsymbol{\xi})) \\ &= \sum_{\boldsymbol{\xi} \in \Omega} \mathbb{P}[\mathbf{Y}, \boldsymbol{\xi} | \boldsymbol{\Theta}]\end{aligned}$$

La méthode usuelle permettant d'estimer les paramètres $\boldsymbol{\Theta} = (\theta_1, \dots, \theta_g, \rho_{local}, \rho_{régional})$ d'une distribution est la méthode du maximum de vraisemblance. Étant donné une réalisation \mathbf{y} de \mathbf{Y} , le maximum de vraisemblance va chercher les paramètres $\hat{\boldsymbol{\Theta}}^{MV}$ maximisant la probabilité d'avoir observé \mathbf{y} sous l'hypothèse que \mathbf{y} a été généré par le modèle proposé :

$$\hat{\boldsymbol{\Theta}}^{MV} = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}} (\mathbb{P}[\mathbf{Y} = \mathbf{y} | \boldsymbol{\Theta}])$$

Cet estimateur bénéficie d'un certain nombre de propriétés : estimateur convergent, asymptotiquement efficace et normalement distribué (voir par exemple la section 2.4 de l'ouvrage de Wood - S. N. WOOD, 2006) ce qui le rend très attractif. Toutefois dans le cas du modèle de mélange, la vraisemblance a une forme non-linéaire coûteuse à évaluer, ce qui rend la maximisation difficile. Aussi d'autres méthodes d'estimation sont utilisées.

L'algorithme EM : cas du modèle de mélange non spatial

L'algorithme classiquement utilisé pour estimer un modèle de mélange est l'algorithme EM (DEMPSTER et al., 1977). Plutôt que de maximiser directement la vraisemblance, l'algorithme EM va maximiser itérativement une borne inférieure de la vraisemblance. L'idée générale de l'algorithme est que, si l'on connaissait la réalisation \mathbf{x} de la variable d'appartenance des sites aux groupes sur notre échantillon, maximiser $\mathbb{P}[\mathbf{Y} = \mathbf{y}, \boldsymbol{\xi} = \mathbf{x} | \boldsymbol{\Theta}]$ est bien plus simple que de maximiser $\mathbb{P}[\mathbf{Y} = \mathbf{y} | \boldsymbol{\Theta}]$.

EXEMPLE : modèle de mélange sans terme spatial ($\rho_{local} = 0$ et $\rho_{régional} = 0$) :

$$\begin{aligned} \operatorname{argmax}_{\boldsymbol{\Theta}} (\mathbb{P}[\mathbf{Y} = \mathbf{y}, \boldsymbol{\xi} = \mathbf{x} | \boldsymbol{\Theta}]) &= \operatorname{argmax}_{\boldsymbol{\Theta}} (\log (\mathbb{P}[\mathbf{Y} = \mathbf{y}, \boldsymbol{\xi} = \mathbf{x} | \boldsymbol{\Theta}])) \\ &= \operatorname{argmax}_{\boldsymbol{\Theta}} \left(\sum_{i=1}^n \log (f_{\theta_{x_i}}(Y_i)) \right) \\ &= \operatorname{argmax}_{\boldsymbol{\Theta}} \left(\sum_{g=1}^G \sum_{i \in \{j | x_j = g\}} \log (f_{\theta_g}(Y_i)) \right) \\ &= \left(\operatorname{argmax}_{\theta_1} \left(\sum_{i \in \{j | x_j = 1\}} \log (f_{\theta_1}(Y_i)) \right), \dots, \operatorname{argmax}_{\theta_G} \left(\sum_{i \in \{j | x_j = G\}} \log (f_{\theta_G}(Y_i)) \right) \right) \end{aligned}$$

On se ramène donc à estimer de façon indépendante les paramètres associés à la distribution de chaque groupe, ce qui dans le cas gaussien revient à évaluer la moyenne et la variance de l'intensité des sites du groupe. La réalisation de la variable aléatoire d'appartenance aux groupes $\boldsymbol{\xi}$ étant inconnue, elle va être approchée par son espérance conditionnelle aux observations, notée $\pi_i^{posterior}$, à l'aide du théorème de Bayes :

$$\mathbb{P}[\xi_i = g | Y_i = y_i, \theta_g] = \frac{\mathbb{P}[\xi_i = g] f_{\theta_g}(Y_i)}{\sum_{\gamma=1}^G \mathbb{P}[\xi_i = \gamma] f_{\theta_\gamma}(Y_i)}$$

Les $\pi_i^{posterior}$ seront appelées par la suite les probabilités d'appartenance a posteriori aux groupes.

De façon générale, l'algorithme EM peut être vu comme une procédure itérative alternant entre deux étapes :

- **L'étape E** (pour espérance) où l'on évalue l'espérance de l'appartenance aux groupes conditionnellement aux observations
- **L'étape M** (pour maximisation) où l'on estime les paramètres $\boldsymbol{\Theta}$ du mélange en maximisant $\mathbb{E}[\mathbb{P}[\mathbf{Y}, \boldsymbol{\xi} | \boldsymbol{\Theta}] | \mathbf{Y}, \widehat{\boldsymbol{\Theta}}]$, où $\widehat{\boldsymbol{\Theta}}$ correspond aux paramètres estimés à l'itération précédente. Cette espérance est généralement appelée vraisemblance complétée.

Cette procédure permet de converger vers un maximum local de la vraisemblance mais n'a aucune garantie de converger vers le maximum global. Elle est donc sensible aux valeurs utilisées pour initialiser la procédure, puisque, suivant ces valeurs, la convergence peut se faire en un maximum local ou vers le maximum global.

L'approximation du champ moyen

L'estimation du modèle de mélange spatial est rendu très difficile à cause de la corrélation spatiale existant entre les observations. En effet, la vraisemblance fait intervenir la distribution spatiale associée au modèle de Potts :

$$\mathbb{P}[\boldsymbol{\xi}|\rho] = \frac{1}{Z} \exp(\rho_{local} V_{local}(\boldsymbol{\xi}) + \rho_{régional} V_{régional}(\boldsymbol{\xi}))$$

où la constante de normalisation du modèle spatial s'exprime comme la somme sur Ω , l'ensemble des configurations spatiales possibles pour les groupes :

$$Z = \sum_{\boldsymbol{\xi} \in \Omega} \exp[\rho_{local} V_{local}(\boldsymbol{\xi}) + \rho_{régional} V_{régional}(\boldsymbol{\xi})]$$

La vraisemblance est en général incalculable puisqu'en considérant G groupes avec n observations, la somme contient alors G^n termes, soit un nombre de termes en $\exp(n)$. Ainsi pour 30 000 voxels et 4 groupes, le nombre de termes de la somme n'est même pas calculable¹.

Pour pouvoir approcher la distribution spatiale, l'approximation du champ moyen a été utilisée. Elle consiste à approximer la vraie distribution de probabilité \mathbb{P} par une autre distribution de probabilité que l'on notera \mathbb{P}^{MF} :

$$\mathbb{P}[\boldsymbol{\xi}] \approx \mathbb{P}^{MF}[\boldsymbol{\xi}]$$


Cette nouvelle distribution de probabilité sera choisie factorisable c'est à dire décomposable en un produit de distributions spécifiques à chaque site :

$$\mathbb{P}^{MF}[\boldsymbol{\xi}] = \prod_{i=1}^n \mathbb{P}_i^{MF}[\xi_i]$$

Pour déterminer la forme de la distribution \mathbb{P}^{MF} , l'approximation du champ moyen va minimiser la divergence de Kullback-Leibler entre \mathbb{P}^{MF} et \mathbb{P} . On montre que, dans le cas d'un modèle spatial limité aux cliques d'ordre 1 et 2, la distribution optimale correspond à la vraie distribution de probabilités où l'on a fixé les variables aléatoires d'appartenance aux groupes des sites voisins ($\boldsymbol{\xi}_{N(i)}$) à une valeur déterministe :

$$\mathbb{P}^{MF}[\boldsymbol{\xi}] = \prod_{i=1}^n \mathbb{P}[\xi_i | \boldsymbol{\xi}_{N(i)} = \overline{\boldsymbol{\xi}_{N(i)}}]$$

Cette valeur, notée $\overline{\boldsymbol{\xi}_{N(i)}}$, est égale à l'espérance de la variable aléatoire. Une démonstration de ce résultat peut être trouvée en annexe 2 de BLANCHET, 2007.

1. Pour donner un ordre d'idée le dernier n pour lequel le nombre de terme est calculable avec le logiciel  est $n = 511$. Il indique 4,494233 10³⁰⁷ termes.

Cette approximation va considérablement simplifier l'estimation puisque le modèle de Potts s'écrit désormais :

$$\begin{aligned}\mathbb{P}^{MF}[\xi|\rho] &= \prod_{i=1}^n \mathbb{P}[\xi_i | \xi_{N(i)} = \overline{\xi_{N(i)}}, \rho] \\ &= \prod_{i=1}^n \frac{1}{Z_i} \exp\left(\rho_{local} V_{local}(\xi_i, \overline{\xi_{N_{local}(i)}}) + \rho_{régional} V_{régional}(\xi_i, \overline{\xi_{N_{régional}(i)}})\right)\end{aligned}$$

avec

$$\begin{aligned}V_{local}(\xi_i, \overline{\xi_{N_{local}(i)}}) &= N_{local}(\xi_i, \overline{\xi_{N_{local}(i)}}) \\ V_{régional}(\xi_i, \overline{\xi_{N_{régional}(i)}}) &= \frac{1}{C_{max}} \sum_{c=1}^{C_{max}} N_c(\xi_i, \overline{\xi_{N_c(i)}}) \\ Z_i &= \sum_{g=1}^G \exp\left(\rho_{local} V_{local}(\xi_i = g, \overline{\xi_{N_{local}(i)}}) + \rho_{régional} V_{régional}(\xi_i = g, \overline{\xi_{N_{régional}(i)}})\right)\end{aligned}$$

Z_i consiste désormais en une somme de G termes ce qui ne pose plus de problème d'évaluation.

L'algorithme EM variationnel

En rendant la distribution de probabilité factorisable, l'approximation du champ moyen modifie l'étape E de l'algorithme EM (estimation des probabilités a posteriori d'appartenance aux groupes). Lors de cette étape, il n'y a plus la garantie d'augmenter la vraisemblance puisque l'optimum peut être une distribution non-factorisable et donc ne pas appartenir à l'espace de recherche. Néanmoins, les propriétés théoriques de l'algorithme EM variationnel s'appliquent ici et garantissent que l'on minimise la divergence de Kullback-Leibler entre la distribution du champ moyen et la vraie distribution à chaque étape de l'algorithme EM (GUNAWARDANA et al., 2005). L'opposé de cette divergence donne une borne inférieure de la vraisemblance qui augmente au cours des itérations.

Nous utiliserons donc un algorithme EM variationnel estimer le modèle de mélange spatial. De façon analogue à un algorithme EM usuel, il consiste en deux étapes :

- **étape E** : elle se divise en trois sous-étapes :
 - Initialisation du vecteur des espérances d'appartenance aux groupes :

$$\overline{\xi_{i,g}} = \frac{f_{\theta_g}(Y_i)}{\sum_{\gamma=1}^G f_{\theta_\gamma}(Y_i)}$$

où $f_{\theta_g}(Y_i)$ est la densité de la distribution du groupe g évaluée en Y_i .

- Résolution d'un problème du point fixe afin d'estimer les probabilités d'appartenance a priori aux groupes (appelées π^{prior} sur la figure) définies par le modèle spatial. Ce point est détaillé dans l'annexe C de l'article page 145.

- Calcul des probabilités d'appartenance a posteriori aux groupes :

$$\pi_{i,g}^{posterior} = \frac{\pi_{i,g}^{prior} f_{\theta_g}(Y_i)}{\sum_{\gamma=1}^G \pi_{i,\gamma}^{prior} f_{\theta_\gamma}(Y_i)}$$

– **étape M** : Elle se divise en deux sous-étapes :

- Estimation des paramètres de la distribution. Pour chaque groupe, un modèle linéaire généralisé pondéré est ajusté. S'il s'agit du groupe g , la pondération se fait avec les $\pi_{i,g}^{posterior}$. Pour une distribution gaussienne où l'on ne spécifie aucune covariable explicative de l'intensité, il suffit de calculer le moyenne et la variance des intensités observées pondérées par les $\pi_{i,g}^{posterior}$.
- Estimation des paramètres spatiaux. Il s'agit de maximiser, par rapport à ρ , le terme associé au modèle de Potts :

$$-\log(Z_i) + \sum_{i=1}^n \sum_{g=1}^G \pi_{i,g}^{posterior} \left(\rho_{local} V_{local}(\xi_i = g | \overline{\xi_{N_{local}(i)}}) + \rho_{régional} V_{régional}(\xi_i = g | \overline{\xi_{N_{régional}(i)}}) \right)$$

où Z_i dépend aussi de ρ_{local} et $\rho_{régional}$. La maximisation est effectuée à l'aide d'une méthode de quasi-Newton appelée "Limited-memory BFGS".

L'ensemble de la procédure est représenté figure [III.13](#).

Spécification du modèle de mélange pour la segmentation de lésions

Le modèle de mélange pour la segmentation de lésion est utilisé avec 4 groupes ainsi que deux variables d'intensité : le T2 FLAIR et le T2 FLAIR après soustraction du controlatéral. Les probabilités d'appartenance au liquide cérébro-spinal, à la substance blanche et à la substance grise (déterminées lors du pré-traitement) sont incluses dans le modèle de structure comme variables explicatives de l'appartenance aux groupes.

En raison d'un nombre important d'artéfacts sur les coupes basses, une étape supplémentaire de correction des probabilités a posteriori est utilisée. Celle-ci consiste à imposer que les voxels appartenant à la lésion soient reliés à des points initiaux appelés graines. Le lien peut être de voisinage direct ou indirect via d'autres voxels appartenant à la lésion. Si ce n'est pas le cas, le voxel est rattaché aux autres groupes proportionnellement à son appartenance initiale. Voir l'annexe [C](#) page [245](#) pour plus de détails.

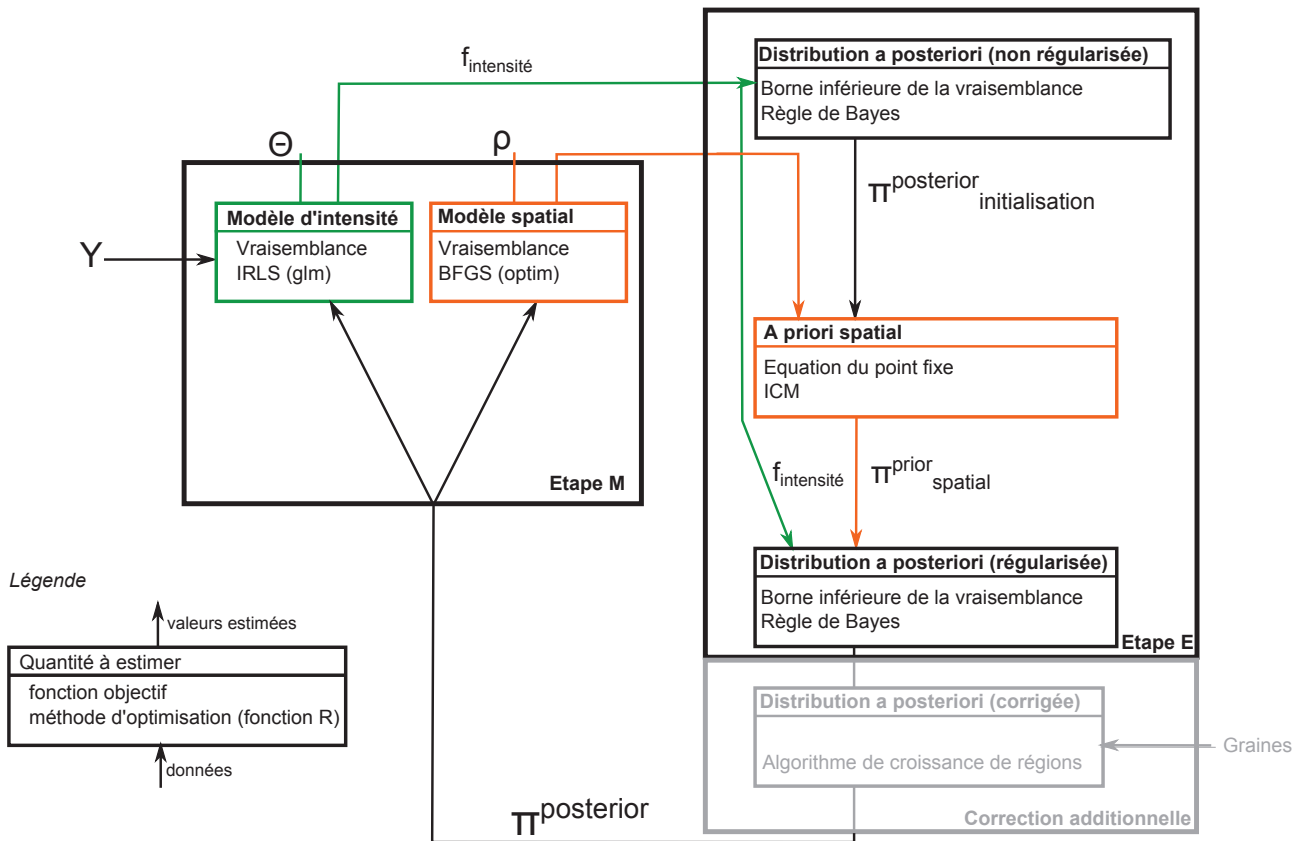


FIGURE III.13 – Schéma récapitulatif des différentes étapes de l'algorithme EM pour l'estimation du modèle de mélange. Les étapes E et M sont itérées jusqu'à stabilisation de la divergence ainsi que des paramètres du modèle. IRLS : moindres carrés pondérés itérés; BFGS : méthode de quasi-Newton; ICM : itération conditionnelle des modes adaptée pour l'approximation du champ moyen (voir l'article de CELEUX et al., 2003, section 4.3)

Principaux résultats

Nous avons testé notre algorithme sur des données simulées ainsi que sur des données réelles. L'étude de simulation a montré que :

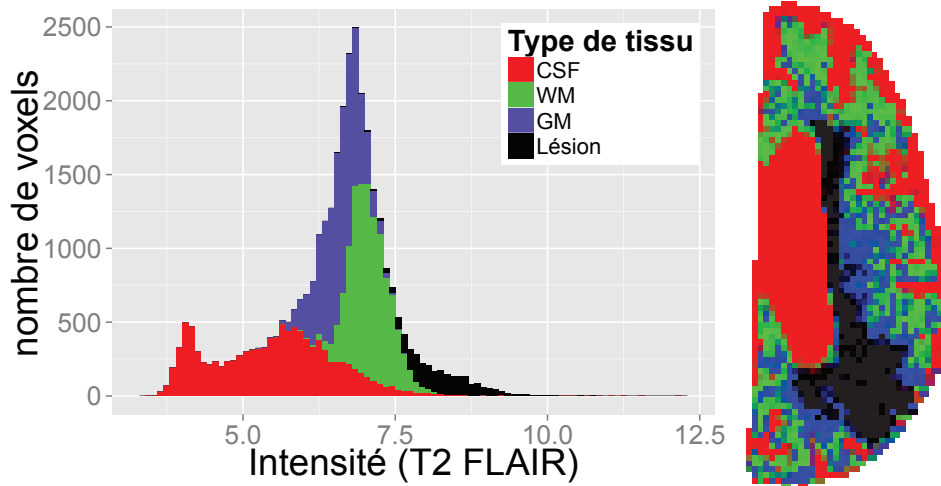
- La régularisation locale est efficace pour traiter un bruit ponctuel mais ne permet pas de corriger des artéfacts même de quelques voxels de large.
- La régularisation régionale permet de corriger les artéfacts. Cependant, plus l'artéfact est intense, plus il nécessite un paramètre de régularisation élevé pour être corrigé.
- L'association de la régularisation locale et régionale donne les meilleurs résultats
- L'approximation du champ moyen induit un biais négligeable pour l'estimation du paramètre spatial local (ρ_1).

L'application de l'algorithme sur données réelles (IRM de patients ayant un AVC) confirme les résultats précédents :

- La régularisation locale améliore la segmentation.
- En présence d'une maladie de la substance blanche, la régularisation régionale permet de limiter ou de supprimer les artéfacts. Dans les autres cas, elle ne semble pas apporter de bénéfice net.

Néanmoins, la méthode admet ses limites dans le cas d'une maladie de la substance blanche de même étendue ou de plus grande étendue que la lésion. En effet, la taille du voisinage sera insuffisante pour permettre la correction de l'artéfact. La correction est aussi limitée dans le cas d'une intensité très élevée de la maladie de la substance blanche. Enfin, le groupe de substance grise obtenu est très hétérogène puisque la plupart des artéfacts sont rattachés à ce groupe lors de la correction (voir figure III.14).

Modèle de mélange non spatial



Modèle de mélange avec régularisation spatiale

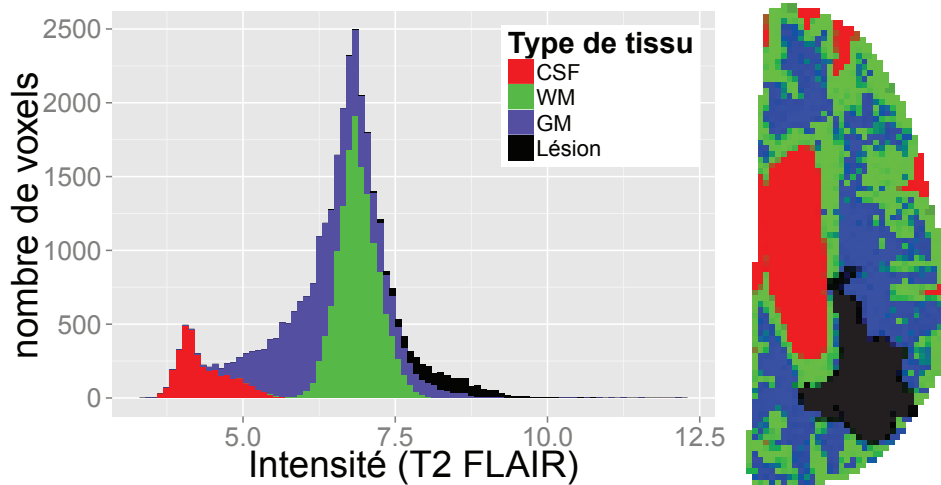


FIGURE III.14 – Groupes formés par le modèle de mélange sans prise en compte de l'information spatiale (1^{ère} ligne) puis en imposant une géométrie régulière des groupes (2nd ligne). La 1^{ère} colonne indique l'histogramme d'intensité de chacun des groupes et la 2nd colonne représente le résultat de la segmentation sur la coupe 15 du patient "009". Les paramètres IRM correspondant sont représentés à la dernière ligne de la figure III.1 page 95.

Article

Le modèle de segmentation que nous proposons est en fait un modèle de mélange formalisant une généralisation du modèle de Potts ainsi que l'utilisation simultanée d'un modèle de structure et d'un modèle spatial pour l'appartenance a priori aux groupes. Pour la publication de ce modèle, un journal méthodologique a été préféré à un journal appliqué. Il nous a semblé utile de montrer les limites de l'estimation automatique des paramètres spatiaux préconisée dans la littérature (WOOLRICH et al., 2005). En effet, la présence d'artéfacts biaise l'estimation des paramètres spatiaux et nuit à la segmentation (figure J du matériel supplémentaire page 158). Nous avons donc proposé une alternative en utilisant les paramètres spatiaux estimés sur des patients sans maladie de la substance blanche. Enfin, à la demande des relecteurs du journal, nous avons comparé notre algorithme à l'algorithme FAST du logiciel FSL, un des logiciels de référence en imagerie médicale.

Biostatistics (2015), pp. 1–16
doi:10.1093/biostatistics/kxv004

Spatially regularized mixture model for lesion segmentation with application to stroke patients

BRICE OZENNE, FABIEN SUBTIL

*Service de Biostatistique, Hospices Civils de Lyon, Lyon, France and Equipe Biostatistique Santé
Université Lyon I, CNRS UMR 5558, 69100, Villeurbanne, France*

LEIF ØSTERGAARD

Center of Functionally Integrative Neuroscience, Århus University, 8000 Århus, Denmark

DELPHINE MAUCORT-BOULCH*

*Service de Biostatistique, Hospices Civils de Lyon, Lyon, France and Equipe Biostatistique Santé
Université Lyon I, CNRS UMR 5558, 69100 Villeurbanne, France*

delphine.maucort-boulch@chu-lyon.fr

SUMMARY

In medical imaging, lesion segmentation (differentiation between lesioned and non-lesioned tissue) is a crucial and difficult task. Automated segmentation algorithms based on intensity analysis have been already proposed and recent developments have shown that integrating spatial information enhances automatic image segmentation. However, spatial modeling is often limited to short-range spatial interactions that deal only with noise or small artifacts. Previous tissue alterations (e.g. white matter disease (WMD)) similar in intensity with the lesion of interest require a broader-scale approach to be corrected. On the other hand, imaging techniques offer now a multiparametric voxel characterization that may help differentiating lesioned from non-lesioned voxels. We developed an unsupervised multivariate segmentation algorithm based on finite mixture modeling that incorporates spatial information. We extended the usual spatial Potts model to the regional scale using a ‘multi-order’ neighborhood potential, with internal adjustment of the regional scale according to the lesion size. We validate the ability of this new algorithm to deal with noise and artifacts (linear and spherical) using artificial data. We then assess its performance on real magnetic resonance imaging brain scans of stroke patients with history of WMD and show that regional regularization was able to remove large-scale WMD artifacts.

Keywords: Expectation-maximization algorithm; Finite mixture models; Image segmentation; Markov random fields; Mean-field approximation.

1. INTRODUCTION

Lesion segmentation provides crucial information for the management of patients with brain tumors, stroke, or multiple sclerosis. Technically, this task aims to differentiate lesion-related from normal voxels

*To whom correspondence should be addressed.

(volumetric picture elements) in an image. In everyday practice, image segmentation is often carried out by visual examination, which is time consuming, source of inter-observer variability, and very difficult in case of 3D images. It would be thus very useful to develop an automated classification algorithm able to identify lesion-related voxels.

The current approaches are mainly based on voxel intensity analysis; they seek the best association between abnormal signals and lesion-related voxels. For example, in T2 magnetic resonance imaging (MRI), a stroke lesion appears as a hyperintense signal. To date, various approaches have been proposed: geometric deformable models (Weinman *and others*, 2003), logistic regression (Sweeney *and others*, 2013), support vector machine (Klöppel *and others*, 2011), K-nearest neighbors (Klöppel *and others*, 2011), fuzzy C-means (Admiraal-Behloul *and others*, 2005), and finite mixture models (fMMs) (Zhang *and others*, 2001; Woolrich *and others*, 2005; Kabir *and others*, 2007; Artzi *and others*, 2013). The latter models have become popular segmentation tools because they do not require a visually segmented reference image and because they offer a flexible modeling framework. Moreover, the combination of several imaging sequences have improved the contrast between normal and abnormal voxels leading to better segmentation (Admiraal-Behloul *and others*, 2005; Klöppel *and others*, 2011).

Today, voxel intensity analyses seem to be insufficient. Previous tissue alterations can have the same intensities as the current lesion making the segmentation task even more challenging (Levy-Cooperman *and others*, 2008). The use of spatial information together with intensity analysis becomes then necessary. Several strategies have been then proposed such as addition of information about the regional intensity using Gaussian smoothing (Sweeney *and others*, 2013) or Gabor filters (Klöppel *and others*, 2011) or incorporation of the spatial correlation into the segmentation model using Potts models (Feng *and others*, 2012; Kabir *and others*, 2007; Woolrich *and others*, 2005; Artzi *and others*, 2013). Potts model, derived from the Markov random field (MRF) theory (Gaetan *and Guyon*, 2000), offers a powerful framework for spatial modeling with a reasonable assumption of local homogeneity (adjacent voxels are likely to belong to the same spatial structure). However, this model considers only short-range neighborhood dependencies. This makes the model robust to noise or tissue heterogeneity but unable to deal with artifacts of more than a few voxels. Indeed, important artifacts are frequently present in MRI brain scans of elderly patients with history of white matter disease (WMD) (Manolio *and others*, 1994) and lead to poor segmentation with tortuous lesion shapes and high false-positive rates. These shapes need to be smoothed or excluded. A regularization tool for large regions is thus clearly necessary.

We present here an unsupervised lesion-segmentation algorithm able to integrate multiparametric data and spatial information and provide a more robust segmentation tool. We combine Hidden Markov Mixture Models (Zhang *and others*, 2001) with concomitant latent class models (Dayton *and Macready*, 1988) and to propose an extension of Potts' model to a regional scale. The aim is providing a flexible algorithm for advanced modeling of MRI intensity signals and spatial regularization.

The present article is organized as follows: Section 2 is a reminder of the common non-spatial mixture model. Section 3 describes the 2 scale-regularization strategies and their integration into the mixture model. Section 4 develops the EM algorithm used to estimate the model parameters. The lesion-segmentation algorithm is tested on artificial data in Section 5 and compared with the physicians' visual segmentation of MRI images of stroke patients in Section 6.

2. THE FMM

Let us consider a regular lattice containing n sites $(z_i)_{i \in \{1, \dots, n\}}$. In a 2D space, these sites would correspond to pixels and, in a 3D space, to voxels. The lattice is divided into G unknown spatial structures $(\mathcal{G}_1, \dots, \mathcal{G}_G)$ we wish to identify. Each site belongs to a specific spatial structure (herein called 'group'), through an indicator variable ξ_i which equals g if site z_i belongs to \mathcal{G}_g . For each site, an intensity Y_i (possibly multi-dimensional) is observed and supposed to enable group identification.

Spatially regularized mixture model for lesion segmentation

3

The fMM models the vector of intensity $\mathbf{Y} = (Y_i)_{i \in \{1, \dots, n\}}$ using a mixture of G distributions. Each group is characterized by a specific distribution of \mathbf{Y} with parameter θ_g . Assuming independence between sites ($\mathcal{H}1$) and denoting $\Theta = \{\theta_1, \dots, \theta_g\}$, the likelihood of the mixture can be expressed:

$$\mathbb{P}[\mathbf{Y}|\Theta] = \prod_{i=1}^n \sum_{g=1}^G \mathbb{P}[Y_i|\xi_i = g, \theta_g] \mathbb{P}[\xi_i = g]$$

where $\mathbb{P}[\xi_i = g]$ denotes the prior probability; i.e. the probability for site z_i to belong to group g before observing the realization of the intensity variable. With no specific information, an uninformative probability distribution (e.g. $1/G$) or a probability distribution that incorporates the knowledge about group attribution (e.g. $\text{card}(\mathcal{G}_g)/n$ where $\text{card}(\mathcal{G}_g)$ denote the cardinality of group g) may be used. Otherwise, a vector of variables $\mathbf{X} = (X_i)_{i \in \{1, \dots, n\}}$ predictive of the group membership is to be considered. In lesion segmentation, various brain regions may have various vulnerabilities to lesion that requires using additional variables. These variables will be called ‘concomitant variables’ and assumed to be independent with \mathbf{Y} conditionally on group membership. Denoting β the parameters associated with \mathbf{X} , the likelihood of the mixture becomes:

$$\mathbb{P}[\mathbf{Y}|\mathbf{X}, \Theta, \beta] = \prod_{i=1}^n \sum_{g=1}^G \mathbb{P}[Y_i|\xi_i = g, \theta_g] \mathbb{P}[\xi_i = g|X_i, \beta] \quad (2.1)$$

Combining information given by \mathbf{X} and \mathbf{Y} allows computing the probability of group membership of each group g , $\pi_g^{\text{posterior}}$. This is done using Bayes theorem:

$$\pi_{i,g}^{\text{posterior}} = \mathbb{P}[\xi_i = g|X_i, Y_i, \Theta] = \frac{\mathbb{P}[Y_i|\xi_i = g, \theta_g] \mathbb{P}[\xi_i = g|X_i, \beta]}{\sum_{\gamma=1}^G \mathbb{P}[Y_i|\xi_i = \gamma, \theta_\gamma] \mathbb{P}[\xi_i = \gamma|X_i, \beta]} \quad (2.2)$$

3. SPATIAL REGULARIZATION

In this section, we will relax hypothesis ($\mathcal{H}1$). Indeed, the independence between spatial position and group membership is unrealistic because the groups are often composed of one or a few spatial clusters of sites. This is why a spatial correlation between neighboring sites is integrated into the fMM to model the spatial homogeneity of the groups. For this, the spatial relationships between the sites have to be defined using a neighborhood structure. For a given site z_i , the closest set of neighbors is called ‘first order neighbors’ and denoted by $\mathbf{N}_1(i)$. Neighborhoods of higher orders c will be denoted by $\mathbf{N}_c(i)$ (for illustration, see Figure 1 in Appendix A of supplementary material available at *Biostatistics* online) and the whole neighborhood of site z_i will be denoted by $\mathbf{N}(i)$.

3.1 Local regularization

Local regularization favors local homogeneity in group membership by using a spatial prior. Potts model is a common specification for the spatial prior that models immediate neighborhood dependencies:

$$\mathbb{P}[\xi] = \frac{1}{Z} \exp \left[\sum_{i=1}^n V_{\text{ext}}(\xi_i) + \rho_1 \sum_{i=1}^n V_{\text{loc}}(\xi_i, \xi_{\mathbf{N}_1(i)}) \right]$$

4

B. OZENNE AND OTHERS

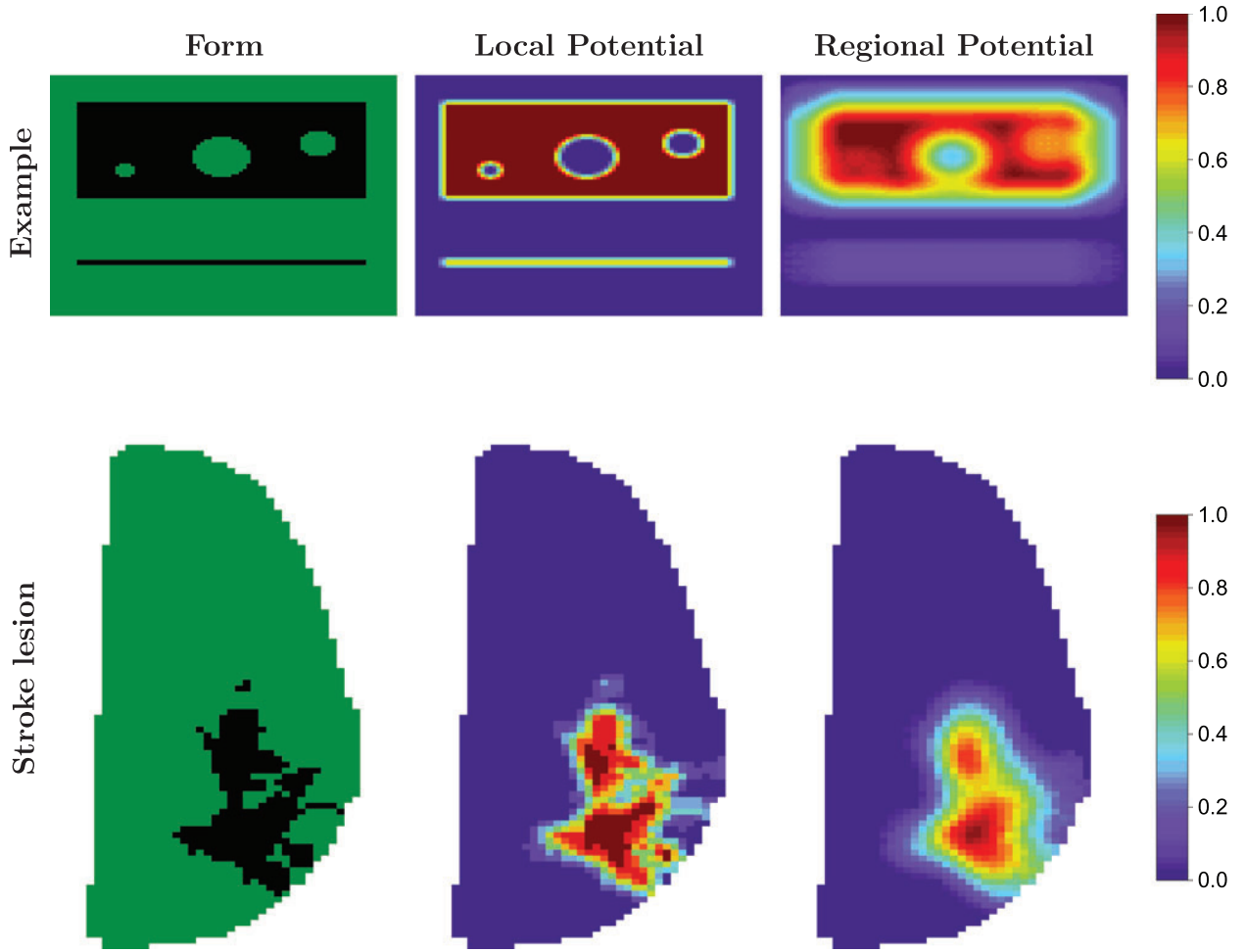


Fig. 1. Local and regional potentials for geometric forms and for an example of stroke lesion. The form is represented in the first column in black, the corresponding local potential is represented in the middle column, and the corresponding regional potential is represented in the right column. The potentials range from 0 [light gray (blue color in online)] to 1 [dark gray (red color in online)]. For the stroke lesion, the potentials were computed with a 3D neighborhood.

Z being a normalizing constant and $\xi_{N_1(i)} = (\xi_j)_{j \in N_1(i)}$. The first sum, called external field potential, incorporates the non-spatial knowledge relative to site group membership. Considering the previous mixture model, this sum would correspond to the log-likelihood of the mixture (2.1). The second sum is a regularization potential that favors a locally homogeneous repartition of sites. Thus, V_{loc} is a homogeneity measure of the group membership: $V_{\text{loc}}(\xi_i, \xi_{N_1(i)}) = \sum_{j \in N_1(i)} w_{i,j} \mathbb{P}[\xi_i = \xi_j]$ where $\mathbb{P}[\xi_i = \xi_j] = \sum_{g=1}^G \mathbb{P}[\xi_j = g] \mathbb{P}[\xi_i = g]$ and $w_{i,j}$ are weights that normalize the site contribution to the potential. $w_{i,j}$ is equal to the inverse of the number of neighbors of site i such that the total weight of each site is one. $V_{\text{loc}}(\xi_i, \xi_{N_1(i)})$ reduces to the usual sum of indicators over neighbor sites $\sum_{j \in N_1(i)} w_{i,j} \mathbb{1}_{\xi_i = \xi_j}$ in the specific case of a binary classification. ρ_1 is a regularization parameter whose minimum value is 0 (when no regularization is applied) and maximum value exceeds rarely 10.

Local regularization is usually performed considering only the immediate neighborhood: 4 or 8 neighbors in 2D images and 6, 18, or 26 neighbors in 3D images. However, because it only considers short-distance spatial dependencies, this approach cannot correct large artifacts (this will be illustrated in the

Spatially regularized mixture model for lesion segmentation

5

Table 1. *Estimated maximum-order neighborhood C_{\max} using the mean distance to the group barycenter for disks of various radiuses and rectangles of several dimensions*

Disk		Rectangle		
Radius	Estimated C_{\max}	Height	Width	Estimated C_{\max}
5	3.1	10	5	4.0
10	6.6	10	10	6.1
15	9.9	10	15	8.3
20	13.3	10	20	10.7
25	16.6	10	25	13.1
30	19.9	10	30	15.5

simulation study, Section 5). In the following section, we propose a new potential that summarizes the information of several neighborhood orders to perform a regional regularization.

3.2 Regional regularization

Regional regularization extends local regularization to a broader scale. By favoring same group memberships between close sites, local regularization tends to smooth locally the form of each group whereas regional regularization favors geometrically regular shapes and penalizes holes, isolated regions, or high curvature forms. For this, regional regularization relies on a regional potential that assesses the neighborhood homogeneity at different distances. This potential is obtained by computing first, for each neighborhood, the proportion of neighbors that belong to the group of the site of interest. Then, the regional potential is the mean of these proportions over all neighborhood orders:

$$V_{\text{reg}}(\xi_i, \xi_{N(i)}) = \frac{1}{C_{\max}} \sum_{c=1}^{C_{\max}} \sum_{j \in N_c(i)} w_{i,j}(c) \mathbb{P}[\xi_j = \xi_i]$$

$w_{i,j}(c)$ normalizes the number of c -order neighbors and C_{\max} is the maximum-order neighborhood. This potential may be seen as an average of local potentials evaluated with various c -order neighbor systems. The maximum-order neighborhood C_{\max} defines the regional range and, thus, should depend on the group dimension; i.e. on the size of the lesion. Indeed, this range may comprise a few voxels in a small lesion but hundreds of voxels in an extended lesion. As a proxy for C_{\max} , we use the mean distance between the voxel position and the group barycenter weighted by the voxel membership to the group. Table 1 shows the estimation of C_{\max} for disks and rectangles of various sizes and Figure 1 displays the values of the local and regional potentials for several forms. In case of fragmented lesions, the calculation of the regional potential takes into account the presence of separated spatial groups, see Appendix F of supplementary material available at *Biostatistics* online for more details.

3.3 Relationship with MRFs

According to the MRF theory, the previous regularization models can be seen as particular cases of a generic probability distribution. Assuming a local Markov property (i.e. the site group membership is only influenced by the neighbor sites), the spatial field can be seen as a discrete MRF. In this case, the

Hammersley Clifford theorem states that the probability distribution of a MRF follows a Gibbs distribution (Besag, 1974).

The probability distribution common to the local and the regional regularization corresponds to a particular Gibbs distribution where the spatial dependencies are restricted to pairs of sites and where the neighborhood system and the potential form are specified. Combining both regularization potentials leads to the following Gibbs distribution:

$$\mathbb{P}[\xi] = \frac{1}{Z} \exp \left[\sum_{i=1}^n V_{\text{ext}}(\xi_i) + \rho_1 \sum_{i=1}^n V_{\text{loc}}(\xi_i, \xi_{N_1(i)}) + \rho_2 \sum_{i=1}^n V_{\text{reg}}(\xi_i, \xi_{N(i)}) \right] \quad (3.1)$$

where $\rho_2 = 0$ corresponds to a pure local regularization and $\rho_1 = 0$ to a pure regional regularization. Denoting $\rho = (\rho_1, \rho_2)$ and considering spatial dependence only for group membership, the likelihood of the mixture can be expressed:

$$\mathbb{P}[\mathbf{Y}|\mathbf{X}, \Theta, \beta, \rho] = \sum_{\mathbf{\Gamma}=(g_1, \dots, g_n) \in [1; G]^n} \mathbb{P}[\mathbf{Y}|\xi = \mathbf{\Gamma}, \Theta] \mathbb{P}[\xi = \mathbf{\Gamma}|\mathbf{X}, \beta, \rho]$$

Because of spatial dependence, the joint prior probability cannot be broken into a product of independent prior probabilities, one per site; this makes the model estimation very tedious. To obtain a more tractable likelihood, we used the mean-field approximation which restores independence between sites by neglecting the fluctuations of neighbor site values (Zhang, 1992). Neighbor sites are thus fixed to their expected value denoted $\bar{\xi}_{N(i)} = (\bar{\xi}_j)_{j \in N(i)}$ leading to the following decomposition: $\mathbb{P}^{\text{MF}}[\xi] = \prod_{i=1}^n \mathbb{P}(\xi_i | \bar{\xi}_{N(i)})$. The mean-field approximation of the likelihood can then be written (see Appendix B of supplementary material available at *Biostatistics* online for the derivation):

$$\mathbb{P}^{\text{MF}}[\mathbf{Y}|\mathbf{X}, \Theta, \beta, \rho] = \prod_{i=1}^n \sum_{g=1}^G \mathbb{P}[Y_i | \xi_i = g, \theta_g] \mathbb{P}[\xi_i = g | X_i, \beta] \mathbb{P}[\xi_i = g | \bar{\xi}_{N(i)}, \rho, V_{\text{ext}} = 0]$$

where the first term corresponds to the contribution of intensity variable, the second term to the contribution of the concomitant variable, and the last term to the contribution of the spatial regularization model to the likelihood:

$$\mathbb{P}[\xi_i = g | \bar{\xi}_{N(i)}, \rho, V_{\text{ext}} = 0] = \frac{1}{Z_i} \exp[\rho_1 V_{\text{loc}}(\xi_i, \bar{\xi}_{N_1(i)}) + \rho_2 V_{\text{reg}}(\xi_i, \bar{\xi}_{N(i)})] \quad (3.2)$$

4. STATISTICAL ESTIMATION

4.1 Variational EM algorithm for spatial fMM

The mean-field approximation is a special case of variational approximation where a joint distribution is approximated by a distribution that factorizes over sites. Its use within the EM framework has been justified by the variational theory and leads to the variational EM algorithm (for an introduction to variational methods, see Jordan and others, 1999, Section 4). Variational EM algorithm maximizes sequentially the lower approximation of the log-likelihood (called lower bound) under the factorization constraint. The maximization is performed according to the factorizing distribution in the E-step and according to the model

parameters in the M-step. The lower bound \mathcal{B} is analogous to the complete log-likelihood of the standard EM algorithm:

$$\mathcal{B}(\Theta, \beta, \rho, \mathbb{P}^{\text{MF}}|\mathbf{Y}, \mathbf{X}) = \sum_{i=1}^n \sum_{g=1}^G \mathbb{P}[\xi_i = g|Y_i, X_i, \bar{\xi}_{\mathbf{N}(i)}] \log(\mathbb{P}[Y_i|\xi_i = g, \Theta]) \quad (4.1a)$$

$$+ \sum_{i=1}^n \sum_{g=1}^G \mathbb{P}[\xi_i = g|Y_i, X_i, \bar{\xi}_{\mathbf{N}(i)}] \log(\mathbb{P}[\xi_i = g|X_i, \beta]) \quad (4.1b)$$

$$+ \sum_{i=1}^n \sum_{g=1}^G \mathbb{P}[\xi_i = g|Y_i, X_i, \bar{\xi}_{\mathbf{N}(i)}] \log(\mathbb{P}[\xi_i = g|\bar{\xi}_{\mathbf{N}(i)}, \rho]) \quad (4.1c)$$

In this formula, the posterior distribution was replaced by its mean-field approximation $\mathbb{P}^{\text{MF}}[\xi|\mathbf{Y}, \mathbf{X}] = \prod_{i=1}^n \mathbb{P}[\xi_i = g|Y_i, X_i, \bar{\xi}_{\mathbf{N}(i)}]$ (for brevity, the dependence on Θ , β , and ρ is omitted). The variational EM algorithm has been shown to converge toward a local minimum of the Kullback–Leiber divergence between the model distribution $\mathbb{P}[\mathbf{Y}, \xi|\mathbf{X}]$ and the mean-field posterior $\mathbb{P}^{\text{MF}}[\xi|\mathbf{Y}, \mathbf{X}]$ under mild conditions. Further details about variational EM and its convergence properties can be found in [Gunawardana and Byrne \(2005\)](#).

4.2 Description of the proposed variational EM algorithm

The proposed segmentation method is performed with an iterative 2-step algorithm (see Appendix C.1 of supplementary material available at *Biostatistics* online for more details on software implementation):

Expectation step: estimation of the mean-field posterior $\mathbb{P}^{\text{MF}}[\xi|\mathbf{Y}, \mathbf{X}]$.

- First, estimation of the spatial prior defined by (3.2) initializing the expected group membership $(\bar{\xi}_i)_{i \in \{1, \dots, n\}}$ by the non-regularized posterior (2.2). Details about the algorithm used for this step can be found in Appendix C.2 of supplementary material available at *Biostatistics* online.
- Second, computing the regularized prior as the product of the concomitant prior $\mathbb{P}[\xi_i = g|X_i, \beta]$ by the spatial prior $\mathbb{P}[\xi_i = g|\bar{\xi}_{\mathbf{N}(i)}, \rho, V_{\text{ext}} = 0]$.
- Finally, computing the (regularized) posterior from (2.2) using the regularized prior.

Maximization step: estimation of the model parameters (Θ, β, ρ) .

- Estimation of the intensity model parameters by maximization of term (4.1a). Assuming independence between the M dimensions of \mathbf{Y} , this maximization is equivalent to that of M independent regressions models weighted by the posterior probabilities.
- Estimation of the concomitant model parameters by maximization of term (4.1b). This maximization is equivalent to that of a multinomial regression weighted by the posterior probabilities.
- Estimation of the regularization parameters by maximization of term (4.1c) is not similar to that of a standard statistical model. Because it is a weighted sum of concave functions, it ensures the uniqueness of the maximum. Thus, it was estimated with a standard optimization routine.

5. EVALUATION ON ARTIFICIAL DATA

5.1 *Data generation and processing*

We tested the robustness of the segmentation model to noise, linear artifacts, and spherical artifacts. We considered 3 groups of 2700 sites each on a regular lattice of 90×90 sites. We generated 3 different datasets on this lattice (see the first row of Figure 2). With the first simulated dataset (Dataset 1), we used the parameterization of [Woolrich and others \(2005\)](#); i.e. a single response variable was simulated in each group according to a normal law with respective means -3 , 0 , and 4 and variances 3 , 1 , and 3 for Groups 1, 2, and 3. Next, the robustness against linear artifacts was tested by incorporating into Group 2 and Group 3 a 2-pixels-wide linear artifact (Dataset 2). The intensity of Group 2 artifact increased linearly from 0.639 to 6.546 . The intensity of Group 3 artifact decreased linearly from 3.361 to -2.546 . Furthermore, the robustness against spherical artifacts was tested by incorporating three disks of radii 2.5 , 5 , and 7.5 pixels into Group 2 and Group 3 (Dataset 3). The intensity of Group 2 artifact ranged from 1.143 to 5.250 with increasing intensities toward the center of the artifact whereas the intensity of Group 3 artifact ranged from -1.250 to 2.857 with decreasing intensities toward the center of the artifact.

We tested 5 specifications of the fMM: no regularization ($\rho = (0, 0)$), local regularization ($\rho = (6, 0)$), strong local regularization ($\rho = (12, 0)$), regional regularization ($\rho = (0, 6)$), and combined local and regional regularization ($\rho = (6, 6)$). By analogy with lesion segmentation, we considered that Groups 1 and 2 represented healthy tissues and Group 3 lesioned tissue. Shape regularity was assumed true only for Group 3 and thus, whenever used, the regional potential was calculated considering Group 3 membership vs. Group 1 and 2 membership. For potential evaluation, the maximum neighborhood-order C_{\max} was arbitrary set at 10, that is, regional dependencies were limited to a span of 10 voxels. Once the model adjusted, each site was assigned to the group with which it had the maximum posterior probability.

Model performance was assessed by the ability of each model to identify correctly Group 3. Two performance metrics were used: the Brier score and the simple matching coefficient. The Brier score is a measure of calibration of predictions computed as the mean squared difference between the simulated Group 3 probability membership and the estimated Group 3 posterior probability. It ranges between 0 (perfect agreement between observation and prediction) and 1 (complete discrepancy). The simple matching coefficient is the ratio of the number of correctly classified sites to the total number of sites. Because artifacts concern only a few sites, bad performance for these sites might not be detected by global performance measures. Thus, for simulated Datasets 2 and 3, we computed also measures of performance restricted to the artifact sites.

5.2 *Results*

Table 2 summarizes the performances of the mixture models with the three simulated datasets. Rows 2–6 of Figure 2 show the Group 3 posterior probabilities estimated by the fMM models. Here, we present simulation results on 2D data but similar results were obtained on 3D data (see Appendix D of supplementary material available at *Biostatistics* online). Mixture models with spatial regularization outperformed clearly the non-spatial mixture model in all 3 simulated datasets using either Brier score or simple matching coefficient. Regarding robustness to noise (Dataset 1), a local regularization was sufficient to deal with most of the heterogeneity in intra-group intensity. The regional regularization was less efficient because of a more important edge effect.

With Dataset 2, 41.3% of the linear artifact could be removed by local regularization ($\rho = (6, 0)$) vs. 23.3% using no regularization. A strong local regularization ($\rho = (12, 0)$) led to better results than a weak one ($\rho = (6, 0)$); i.e. 68.7% artifact removal. However, with equal regularization intensity, regional regularization ($\rho = (0, 6)$) gave better results (60.0% vs. 41.3%) than local regularization ($\rho = (6, 0)$) despite the persistence of residual noise. Local plus regional regularization performed the best with nearly no residual

Spatially regularized mixture model for lesion segmentation

9

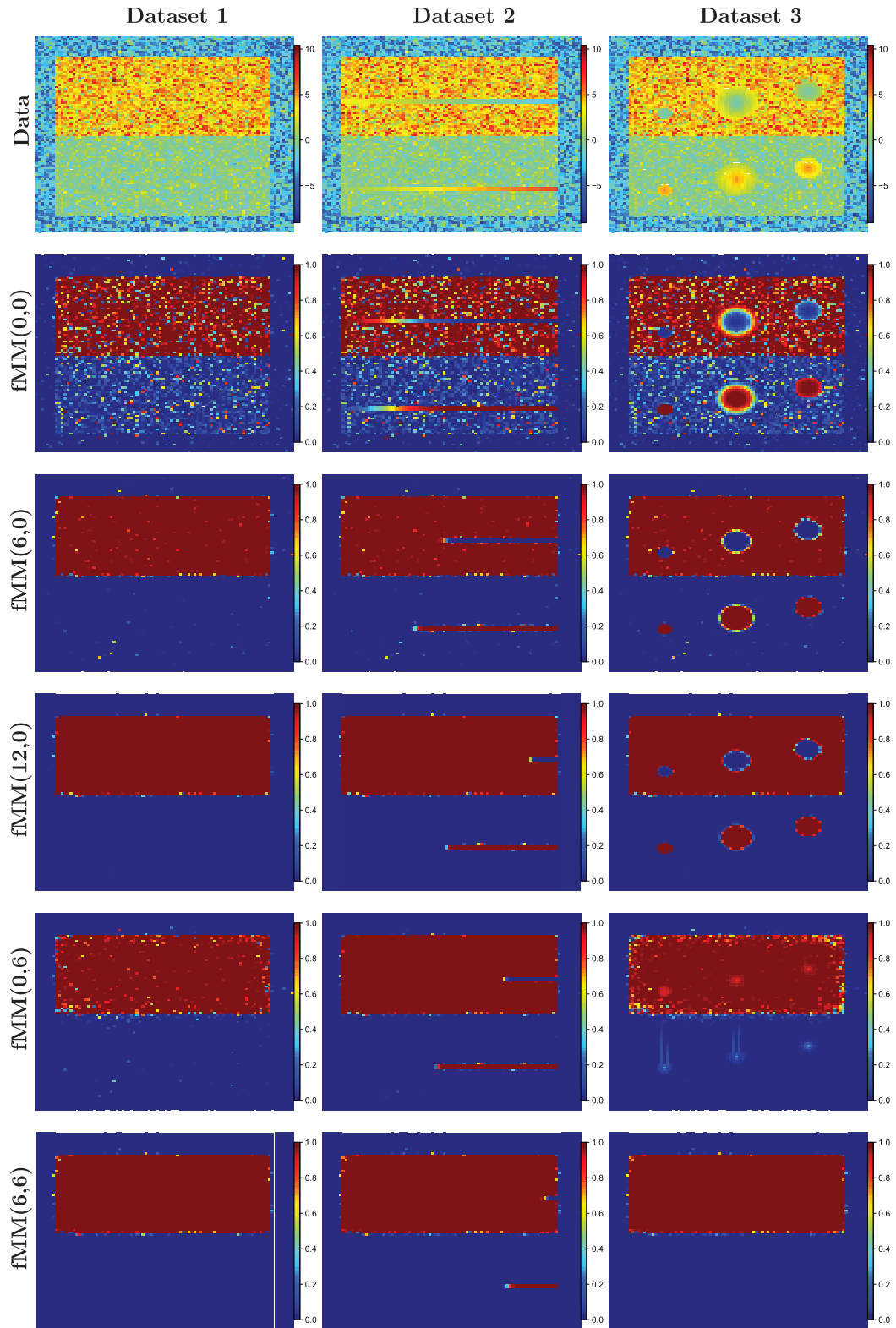


Fig. 2. Posterior probabilities of Group 3 membership with the 5 specifications of the fMM and the 3 simulated datasets.

Table 2. *Evaluation of Group 3 classification with the fMM and different specifications of the spatial regularization parameter ρ*

Dataset	Performances	$\rho = (0,0)$	$\rho = (6,0)$	$\rho = (12,0)$	$\rho = (0,6)$	$\rho = (6,6)$
		Brier score (%)				
Dataset 1	Global	18.8	3.91	2.55	6.54	3.37
Dataset 2	Global	24.3	15.4	11.1	12.5	7.78
Dataset 2	Artifact sites	83.0	76.1	56.1	63.1	36.9
Dataset 3	Global	26.6	21.1	20.2	7.04	3.17
Dataset 3	Artifact sites	76.8	80.4	77.9	4.15	0.03
Simple matching coefficient						
Dataset 1	Global	95.4	99.8	99.9	99.5	99.9
Dataset 2	Global	92.9	97.6	98.7	98.4	99.4
Dataset 2	Artifact sites	23.3	41.3	68.7	60.0	86.0
Dataset 3	Global	91.1	95.3	95.7	99.4	99.9
Dataset 3	Artifact sites	27.0	32.8	36.9	100.0	100.0

Dataset 1 was used to test the robustness to heterogeneity in group intensity. Dataset 2 was used to test the robustness to linear artifacts. Dataset 3 was used to test the robustness to spherical artifacts.

noise left and removal of most of the artifacts (86.0%). With Dataset 3, local regularization did not succeed in removing spherical artifacts; there was only a very slight improvement vs. the non-regularized model (32.8% vs. 27.0%) even with a strong regularization (36.9%). However, the regional regularization gave very good results and was able to remove entirely the artifact.

6. APPLICATION TO SEGMENTATION OF STROKE LESIONS

6.1 Data

We considered 12 patients with unihemispheric stroke lesions in subacute stage (1 month after the onset of symptoms) from the I-know study. This sample was representative of the diversity of stroke configurations as seen on T2 FLAIR images: Patients 1–4 had T2 hyperintense stroke lesions, Patients 5–8 had heterogeneous T2 hyperintense stroke lesions, and Patients 9–12 had WMD close to or overlapping the stroke lesions. Patients 1–4 were used to estimate the regularization parameters and Patients 5–12 to test the relevance of spatial regularization in lesion segmentation.

Imaging data were acquired with a 1.5T MR imaging system (20 slices of 128×128 voxels and voxel size $1.875 \times 1.875 \times 6$ mm). The only pre-processing operation was image co-registration. A probabilistic white matter (WM) / gray matter (GM), cerebrospinal fluid (CSF) segmentation was obtained from T1 sequences. Visual segmentation was performed by three physicians using T2 Fluid-Attenuated-Inversion-Recovery sequence (T2 FLAIR), the gold standard MRI sequence for identification of stroke lesions in subacute stage. Each voxel was included in the final mask when deemed lesioned by 2 of the 3 physicians.

6.2 Model specification

We compared three mixture-model specifications: non-spatial (fMM_{ns}), locally regularized (fMM_{reg1}), and locally and regionally regularized (fMM_{reg2}). For computational reasons, the regional potential span was limited to 10 mm (i.e. 210 neighbors). We modeled the distribution of voxel intensity using a 4-group mixture model with Gaussian distributions. In order to use the symmetry of WMD, we created from the

Spatially regularized mixture model for lesion segmentation

11

Table 3. *Agreements between the physicians' mask and the non-spatial (fMM_{ns}), the local (fMM_{reg1}), and the regional (fMM_{reg2}) mixture models*

$\hat{\rho}$	V_{lesion} (mL)	fMM_{ns} (0, 0)	fMM_{reg1} (4.59, 0)	fMM_{reg2} (3.85, 2.61)	fMM_{ns} (0, 0)	fMM_{reg1} (4.59, 0)	fMM_{reg2} (3.85, 2.61)
		Calibration (%)			Brier score		
Patient 5	45	1.180	0.799	0.687	15.9	17.5	18.7
Patient 6	59	0.935	0.975	0.975	12.7	11.2	11.3
Patient 7	113	0.879	0.933	0.932	20.2	17.3	17.1
Patient 8	49	0.903	0.905	0.887	11.6	11.4	11.4
Patient 9	30	1.62	1.68	1.17	17.1	17.8	13.7
Patient 10	6	2.76	1.16	1.10	11.8	5.41	5.22
Patient 11	11	4.85	1.41	1.29	21.2	9.68	9.60
Patient 12	26	1.44	1.32	1.02	14.9	13.8	13.0
		Simple matching coefficient (%)			Jaccard coefficient (%)		
Patient 5		96.8	96.8	96.4	68.2	62.6	57.0
Patient 6		98.2	98.6	98.6	79.6	84.6	84.6
Patient 7		95.6	96.8	96.9	75.0	81.7	82.3
Patient 8		98.5	98.6	98.6	80.8	82.3	81.9
Patient 9		96.4	96.3	97.9	51.9	52.0	63.0
Patient 10		98.0	99.7	99.7	32.1	74.2	75.2
Patient 11		93.2	99.0	99.0	18.8	59.9	58.7
Patient 12		97.2	97.8	98.2	50.2	57.3	58.7

The agreement was measured with 4 indexes: Brier score, global calibration, simple matching coefficient, and Jaccard coefficient. The optimal values are, respectively, 0, 1, 100% and 100%. The first column of the table indicates the lesion volume in mL.

T2 FLAIR signal an additional parameter called ‘T2 FLAIR contro’; this parameter was computed as the difference in T2 FLAIR intensity between a voxel and its contralateral counterpart. We then adopted a bivariate response model with *T2 FLAIR* and *T2 FLAIR contro*. In addition, a concomitant model was used; it included the membership probabilities of *WM*, *GM*, and *CSF* as predictors. As T2 data present high-intensity non-lesioned cerebral structures, we considered an additional step in our EM algorithm. The stroke location was previously indicated by a physician, with 1 voxel reference point for each lesioned spatial group. At the end of each E-step, we formed lesioned groups using a growing region algorithm (Revol and Jourlin, 1997) with reference points for seeds. Then, voxels located out of the lesioned group had their lesion-group posterior set to 0.

Two strategies were investigated for estimating the regularization parameters: (i) *internal estimation* where the parameters are estimated at each M-step of the EM algorithm, as described in Section 4.2. (ii) *External estimation* using a sample of patients without WMD or intensity heterogeneity. For this, the non-spatial lesion-segmentation algorithm was applied to Patients 1–4. For each segmentation, the regularization parameters (ρ) were estimated by maximizing the probability distribution associated with the regularization model (term (4.1c)). Each estimated parameter was the average of the 4 values relative to the 4 patients.

6.3 *Comparison between computer-aided and visual segmentation*

To test the quality of the three mixture model segmentations, we looked for the agreement between each automated segmentation and the physicians' segmentation and added to the previous 2 metrics (Brier score

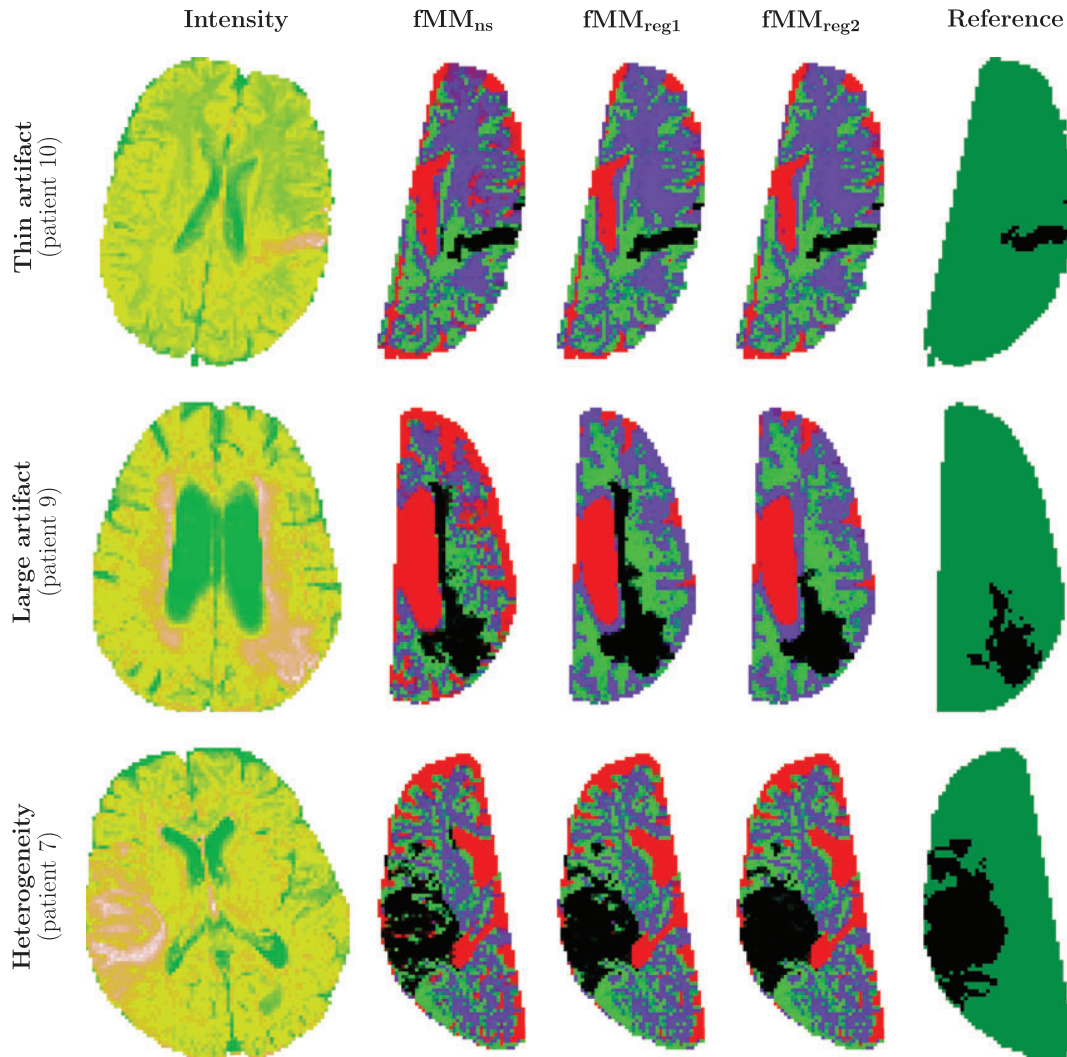


Fig. 3. Results of the mixture models applied to Patients 7, 9, and 10. From left to right: T2 FLAIR cartography, estimated membership probability for the 3 mixture models [lesion in dark gray (CSF in red, WM in blue, GM in green, lesion in black for online)] and the corresponding physicians' mask [lesioned tissue in dark gray, healthy tissue in light gray (lesioned tissue in black, healthy tissue in green for online)].

and simple matching coefficient) a global calibration and Jaccard coefficient. Global calibration is a measure of the accuracy of lesion volume estimation, which is of great clinical interest. It is computed as the ratio of the estimated volume to the true volume. Jaccard coefficient is a measure of similarity between 2 sample sets; it is preferred to the simple matching coefficient in case of data with imbalanced groups ([Bouix and others, 2007](#)). Regarding simulated Group 3 and estimated Group 3, this coefficient was computed as the ratio of the size of the intersection of the 2 sets to the size of their union.

6.4 Results

The comparison between the internal and the external estimation of the regularization parameters is reported in Appendix E.2 of supplementary material available at *Biostatistics* online. Both methods tended to over-estimate the local regularization parameter and underestimate the regional regularization

parameter. However, much larger discrepancies were found with the internal estimation in comparison with the external estimation. Thus, only the external estimation was retained.

Table 3 summarizes the quality of lesion segmentation in the 8 patients with the 3 fMM specifications. Brier score, simple matching coefficient, and Jaccard coefficient led to similar conclusions. A visual assessment of the segmentation quality is shown in Figure 3 for 3 patients. Overall, the non-spatial mixture model was influenced by the heterogeneity in lesion intensity; this led to small underestimations of the volumes (from -6.5% to -12.1% for Patients 6–8) due to holes in the segmentation or confusion between healthy and lesioned voxels (Patient 5). Nevertheless, Jaccard coefficient was relatively high; it ranged from 68.2 to 80.8. WMD influenced more dramatically the quality of the segmentation; this led to much lower Jaccard coefficients (from 18.8 to 51.9) and clear overestimations of the volumes (from $+44.0\%$ to $+4850\%$).

A local regularization with an estimated parameter of 4.59 improved the segmentation in all but 1 patient (Patient 5). It removed small artifacts and smoothed the lesion shapes but enhanced WM artifacts (see Figure 3). In patients with WMD the segmented volume was still overestimated (from $+16\%$ to $+68\%$), especially in Patient 9 ($+68\%$ compared with $+62\%$ with the non-regularized model) because of an extended WMD. Regional and local regularization with parameters estimated at $\rho = (3.85, 2.61)$, respectively, removed at least partially the WM artifacts. It resulted in a better calibration in all patients with WMD despite the possibility of further improvement of segmentation in Patient 11. However, this did not improve segmentation in patients with heterogeneity; the results were very close to those obtained with local regularization alone. We should note here that both regularizations failed in Patient 5 because of a too strong heterogeneity and a very irregular lesion shape (see Appendix G of supplementary material available at *Biostatistics* online).

We investigated the influence of pre-processing on segmentation results (results not shown). Filtering the intensity parameter with a Gaussian kernel led to poor segmentations in all model specifications, probably because of edge blurring. Using an edge-preserving filter (a Gaussian kernel weighted by the difference in intensity between the site and its neighbors) gave satisfactory results in patients with intensity heterogeneity but not in patients with WMD. Finally, we included a concomitant model into the previous segmentation models to exploit the T1 MRI sequence. Withdrawing this model led to close performances but with faster computations.

7. DISCUSSION

We present here a new lesion-segmentation procedure that proved to be robust to noise and intensity artifacts. This procedure is based on a fMM, a flexible model that can include various *a priori* risks for the latent groups using a multinomial model. The intensity model uses Gaussian and Gamma distributions of the latent group response but can be extended to other members of the exponential family. Assuming independence between response variables, it allows multivariate response modeling with possible inclusion of covariates.

The present results confirm the interest of spatial regularization in lesion segmentation. Potts model appears to be very efficient though it is prone to edge-oversmoothing and has to be modified to deal with specific geometries (e.g. [Gottardo and others, 2006](#)). In our application, we wanted to eliminate from the segmented lesion some parts characterized by elongated shapes and fill unrealistic holes inside huge lesions. Because lesion shapes are irregular and vary from one patient to another, a deterministic geometric criterion would not be appropriate. By extending Potts model to regional regularization, we obtained satisfactory results despite longer times required by the computation of the regional potentials (15 vs. 11 min). We compared our algorithm with FAST tool of FSL software (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FAST>). The underlying model of FAST is equivalent to our fMM but with only local regularization and without

concomitant model or region-growing step. Compared with our fMM with only local regularization, FAST gave similar segmentation results though it was less robust to artifacts, which is due to the absence of the region-growing step (see Appendix H of supplementary material available at *Biostatistics* online).

The present simulations have shown that local and regional regularizations are complementary approaches: the former deals with noise or heterogeneous intensity whereas the latter corrects high-intensity artifacts with tortuous shapes. Yet, the regional regularization has some limitations: it cannot deal with very intense or very large artifacts and may oversmooth the lesion shape, especially in the case of large lesions with tortuous shapes. In absence of artifacts, it gives close results to those of local regularization but requires more computer time. Thus, to obtain the best segmentation, one should use local and regional regularization when intensity artifacts are close to the lesion. Otherwise, local regularization is sufficient. One limitation of our approach is that the EM algorithm is known to be sensitive to the initialization step. Thus, for robustness against local optima, one should use several initializations and keep the best segmentation. Besides, we used k -means centers for good initializations.

Estimating the regularization parameters in a missing data context is a difficult because of 2 possible sources of bias. The first source is due to the absence of accurate and tractable estimator for large data. By using the mean-field approximation, we obtained a pseudo-likelihood estimator which has good theoretical properties and was computationally efficient. In the Gaussian case, this estimator is known to be consistent, asymptotically normal, but less efficient than the maximum likelihood estimator (Gaetan and Guyon, 2000). Moreover, our simulation study (see Appendix E.1 of supplementary material available at *Biostatistics* online) shows limited bias in case of large samples, which is often the case with MRI data. Exact approaches have been recently proposed for estimating the local regularization parameter (e.g. Pereyra and others, 2013) but, to be applicable, they should be extended to the regional approach. The second source of bias stems from the use of the estimated group membership instead of the true, but unknown, group membership. As shown in Appendix E.2 of supplementary material available at *Biostatistics* online, this precluded the use of an adaptive estimation as recommended by Woolrich and others (2005). Indeed, the presence of large-scale artifacts led to underestimate the regional regularization parameter making these artifacts difficult to correct. To avoid this, we estimated first the spatial regularization parameters using examples of correctly segmented lesions. This approach has been successfully used in other works (Feng and others, 2012; Gottardo and others, 2006) though it assumes a common geometry of the lesions in all patients. A more realistic approach would be to use it as a prior distribution of the regularization parameter in conjunction with the adaptive estimation. This would adapt the regularization to each individual lesion geometry and enhance the segmentation.

8. SOFTWARE

Software is available upon request. Please contact delphine.maucort-boulch@chu-lyon.fr.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at <http://biostatistics.oxfordjournals.org>.

ACKNOWLEDGMENTS

We are grateful to the I-know consortium for providing the image data. *Conflict of Interest*: None declared.

Spatially regularized mixture model for lesion segmentation

15

FUNDING

The research is supported by a grant from the French Ministry of Higher Education and Research.

REFERENCES

- ADMIRAAL-BEHOUL, F., VAN DEN, H., VAN OSCH, M. J., VAN DER, G. AND REIBER, J. H. (2005). Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. *NeuroImage* **28**(3), 607–617.
- ARTZI, M., AIZENSTEIN, O., JONAS-KIMCHI, T., MYERS, V., HALLEVI, H. AND BASHAT, D. B. (2013). FLAIR lesion segmentation: application in patients with brain tumors and acute ischemic stroke. *European Journal of Radiology* **82**(9), 1512–1518.
- BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Serie B (Methodological)* **36**(2), 192–236.
- BOUIX, S., MARTIN-FERNANDEZ, M., UNGAR, L., NAKAMURA, M., KOO, M. S., MCCARLEY, R. W. AND SHENTON, M. E. (2007). On evaluating brain tissue classifiers without a ground truth. *NeuroImage* **36**(4), 1207–1224.
- DAYTON, C. M. AND MACREADY, G. B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association* **83**(1), 173–178.
- FENG, D., TIERNEY, L. AND MAGNOTTA, V. (2012). MRI tissue classification using high-resolution Bayesian hidden Markov normal mixture models. *Journal of the American Statistical Association* **107**(497), 102–119.
- GAETAN, C. AND GUYON, X. (2000) *Spatial Statistics and Modeling*. Berlin: Springer.
- GOTTARDO, R., BESAG, J., STEPHENS, M. AND MURUA, A. (2006). Probabilistic segmentation and intensity estimation for microarray images. *Biostatistics* **7**(1), 85–99.
- GUNAWARDANA, A. AND BYRNE, W. (2005). Convergence theorems for generalized alternating minimization procedures. *Journal of Machine Learning Research* **6**, 2049–2073.
- JORDAN, M. I., GHAHRAMANI, Z., JAAKKOLA, T. S. AND SAUL, L. (1999). An introduction to variational methods for graphical models. *Machine Learning* **37**, 183–233.
- KABIR, Y., DOJAT, M., SCHERRER, B., FORBES, F. AND GARBAY, C. (2007). Multimodal MRI segmentation of ischemic stroke lesions. *29th Annual International Conference of the IEEE* **1**, 1595–1598.
- KLÖPPEL, S., ABDULKADIR, A., HADJIDEMETRIOU, S., ISSLEIB, S., FRINGS, L., THANH, H. N., MADER, I., TEIPEL, S. J., HÜLL, M. AND RONNEBERGER, O. (2011). A comparison of different automated methods for the detection of white matter lesions in MRI data. *NeuroImage* **57**(2), 416–422.
- LEVY-COOPERMAN, N., RAMIREZ, J., LOBAUGH, N. J. AND BLACK, S. E. (2008). Misclassified tissue volumes in Alzheimer disease patients with white matter hyperintensities: importance of lesion segmentation procedures for volumetric analysis. *Stroke* **39**(4), 1134–1141.
- MANOLIO, T. A., KRONMAL, R. A., BURKE, G. L., POIRIER, V., O'LEARY, D. H., GARDIN, J. M., FRIED, L. P., STEINBERG, E. P. AND BRYAN, R. N. (1994). Magnetic resonance abnormalities and cardiovascular disease in older adults. *Stroke* **25**(2), 318–27.
- PEREYRA, M., DOBIGEON, N., BATATIA, H. AND TOURNERET, J. Y. (2013). Estimating the granularity coefficient of a Potts–Markov random field within an MCMC algorithm. *IEEE Transactions on Image Processing* **22**(6), 2385–2397.
- REVOL, C. AND JOURLIN, M. (1997). A new minimum variance region growing algorithm for image segmentation. *Pattern Recognition Letters* **18**(3), 249–258.

- SWEENEY, E. M., SHINOHARA, R. T., SHIEE, N., MATEEN, F. J., CHUDGAR, A. A., CUZZOCREO, J. L., CALABRESI, P. A., PHAM, D. L., REICH, D. S. AND CRAINICEANU, C. M. (2013). Oasis is automated statistical inference for segmentation, with applications to multiple sclerosis lesion segmentation in MRI. *Neuroimage Clinical* **2**, 402–13.
- WEINMAN, J., BISSIAS, G., HOROWITZ, J., RISEMAN, E. AND HANSON, A. (2003). Nonlinear diffusion scale-space and fast marching level sets for segmentation of MR imagery and volume estimation of stroke lesions. *MICCAI* **2879**, 496–504.
- WOOLRICH, M. W., BEHRENS, T. E. J., BECKMANN, C. F. AND SMITH, S. M. (2005). Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data. *IEEE Transactions on Medical Imaging* **24**(1), 1–11.
- ZHANG, J. (1992). The mean-field theory in EM procedures for Markov random fields. *IEEE Transactions on Signal Processing* **40**(10), 2570–2583.
- ZHANG, Y., BRADY, M. AND SMITH, S. (2001). Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging* **20**(1), 45–57.

[Received July 16, 2014; revised January 20, 2015; accepted for publication January 22, 2015]

APPENDIX

A. NEIGHBORHOOD SYSTEM

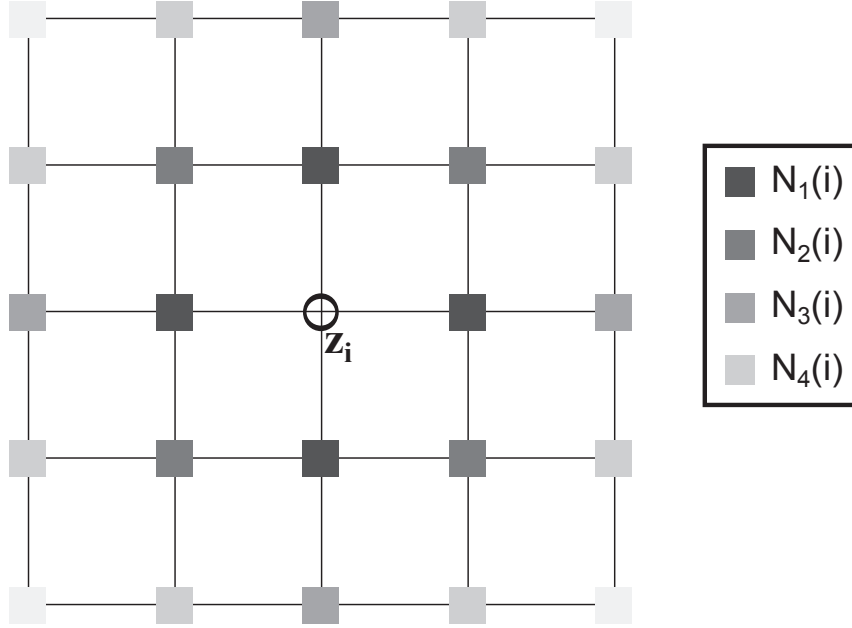


Fig. A. Representation of the first four neighborhood orders concerning the central site of a 5×5 regular two-dimensional lattice. Each shade of grey marks the membership to a given order neighborhood. In this case, $C_{max} = 4$ and $N(i) = \bigcup_{c=1}^4 N_c(i)$

B. MEAN-FIELD APPROXIMATION FOR SPATIAL FINITE MIXTURE MODELS

B.1 *Mean-Field Theory*

Mean-field approximation arises from statistical mechanics Chandler (1987) where it was used to study interparticle interactions. It has been adapted for the EM procedure in image segmentation by Zhang (1992) and is now widely spread (e.g. Hofmann and Buhmann (1997); Celeux *and others* (2003); Kabir *and others* (2007)). It is used to approximate an intractable joint probability distribution \mathbb{P} by a factorizing probability distribution that we will denote \mathbb{P}^{MF} . For instance when considering a random field $\boldsymbol{\xi} = (\xi_i)_{i \in \{1, \dots, n\}}$ it enables to write the probability distribution

as a product over sites $\mathbb{P}^{MF}[\xi] = \prod_{i=1}^n \mathbb{P}^{MF}[\xi_i]$ which may simplify the computations.

By constraining the probability distribution to factorize, mean-field approximation neglects interactions between neighboring sites ξ_j and fix their value to their expectation $\overline{\xi_j}$. It leads to the following approximation (Celeux *and others* (2003)):

$$\mathbb{P}[\xi] \approx \mathbb{P}^{MF}[\xi] = \prod_{i=1}^n \mathbb{P}[\xi_i | \overline{\xi_{N(i)}}]$$

Mean-field approximation can thus be seen as a variational approximation since it replaces a complex function by a simpler one by a decoupling of the degrees of freedom. Finally, mean-field approximation has been shown to be the optimal approximation over the distributions that factorize first relatively to the Gibbs-Bogoliubov-Feynman bound (Chandler (1987)) and then in the sense of the Kullback-Leiber divergence between the true and the factorizing distribution (Hofmann and Buhmann (1997)).

B.2 Variational EM-algorithm

EM algorithm is a maximum likelihood method used to solve intractable optimization problems by introducing additional variables, usually called latent variable. It is especially well suited to image segmentation because these latent variables link the voxels to the groups that are to identify. Standard EM algorithm iterates between an E-step where it seeks the posterior distribution giving the best classification and a M-step where it estimates the group parameters. However, introduction of spatial dependency makes both steps intractable. Indeed the computation of the normalizing constant of spatial probability distribution (denoted Z in the article) require to sum over all the possible spatial configurations of the group membership which is impossible except for very small spatial fields.

Variational EM algorithm has been developed to allow approximating E-steps whenever the best distribution is searched in a restrictive space of distributions. The EM algorithm with the

mean-field approximation is thus a special case of Variational EM algorithm in which the distribution space is restricted to the distributions that factorize. As a consequence, the optimal distribution may not match with the posterior distribution because the posterior distribution may not belong to the restrictive distribution space. Thus, the likelihood might not increase over iterations. Nevertheless, Variational EM algorithm minimizes sequentially the Kullback-Leiber divergence between the model distribution \mathbb{P} (the one that assigns the highest probability to the observed data) and the factorizing distribution \mathbb{P}^{MF} , according to the \mathbb{P}^{MF} distribution during the E-step and according to the model parameters during the M-step (Gunawardana and Byrne (2005)). The opposite of the divergence is called lower bound because it gives a lower estimate than the log-likelihood and increases over iterations. This procedure has been showed to converge toward local minima of the divergence under mild conditions (the variational E-step has to have a unique minimizer (Gunawardana and Byrne (2005)). When using the mean-field approximation, these minima match with the local minima of the likelihood function only if there are stationary points of the variational error function.

Note that the standard EM algorithm (Dempster *and others* (1977)) is a specific case of the Variational EM algorithm with no constrain on the distribution space. Then, the optimal distribution obtained at each E-step is the posterior distribution for which the lower bound reaches the log-likelihood. In this context, the lower bound is usually called completed log-likelihood.

B.3 Derivation of the likelihood under mean-field approximation

The likelihood of a spatial finite mixture model that assumes independence between intensity observations conditionally to group membership but spatial dependence between group membership can be expressed as follows:

$$\mathbb{P}[\mathbf{Y}|\mathbf{X}, \Theta, \beta, \rho] = \sum_{\mathbf{\Gamma}=(g_1, \dots, g_n) \in [1;G]^n} \mathbb{P}[Y, \boldsymbol{\xi} = \mathbf{\Gamma}|\mathbf{X}, \Theta, \beta, \rho]$$

$$\begin{aligned}
&= \sum_{\mathbf{\Gamma}=(g_1, \dots, g_n) \in [1;G]^n} \mathbb{P}[\mathbf{Y}|\mathbf{\xi} = \mathbf{\Gamma}, \Theta] \mathbb{P}[\mathbf{\xi} = \mathbf{\Gamma}|\mathbf{X}, \rho, \beta] \\
&= \sum_{\mathbf{\Gamma}=(g_1, \dots, g_n) \in [1;G]^n} \left(\prod_{i=1}^n \mathbb{P}[Y_i|\xi_i = g_i, \theta_{g_i}] \right) \mathbb{P}[\mathbf{\xi} = \mathbf{\Gamma}|\mathbf{X}, \beta, \rho]
\end{aligned}$$

where the sum over $\mathbf{\Gamma} = (g_1, \dots, g_n) \in [1;G]^n$ corresponds to the sum over all possible spatial configuration for the site group membership.

Introducing mean-field approximation into the likelihood gives the following expressions:

$$\begin{aligned}
\mathbb{P}^{MF}[\mathbf{Y}|\mathbf{X}, \Theta, \beta, \rho] &= \sum_{\mathbf{\Gamma}=(g_1, \dots, g_n) \in [1;G]^n} \left(\prod_{i=1}^n \mathbb{P}[Y_i|\xi_i = g_i, \theta_{g_i}] \right) \left(\prod_{i=1}^n \mathbb{P}[\xi_i = g_i|X_i, \overline{\mathbf{\xi}_{N(i)}}, \beta, \rho] \right) \\
&= \sum_{\mathbf{\Gamma}=(g_1, \dots, g_n) \in [1;G]^n} \prod_{i=1}^n \mathbb{P}[Y_i|\xi_i = g_i, \theta_{g_i}] \mathbb{P}[\xi_i = g_i|X_i, \overline{\mathbf{\xi}_{N(i)}}, \beta, \rho] \\
&= \prod_{i=1}^n \sum_{g=1}^G \mathbb{P}[Y_i|\xi_i = g, \theta_g] \mathbb{P}[\xi_i = g|X_i, \overline{\mathbf{\xi}_{N(i)}}, \beta, \rho]
\end{aligned}$$

Then, considering the previous decomposition with Bayes rule and the independence between $\overline{\mathbf{\xi}_{N(i)}}$ and X_i , it follows:



$$\begin{aligned}
\mathbb{P}[\xi_i = g|X_i, \overline{\mathbf{\xi}_{N(i)}}, \beta, \rho] &= \frac{\mathbb{P}[X_i, \overline{\mathbf{\xi}_{N(i)}}|\xi_i = g, \beta, \rho] \mathbb{P}[\xi_i = g|\beta, \rho]}{\mathbb{P}[X_i, \overline{\mathbf{\xi}_{N(i)}}|\beta, \rho]} \\
&= \frac{\mathbb{P}[X_i|\xi_i = g, \beta, \rho] \mathbb{P}[\overline{\mathbf{\xi}_{N(i)}}|\xi_i = g, \beta, \rho] \mathbb{P}[\xi_i = g|\beta, \rho]}{\mathbb{P}[X_i|\beta, \rho] \mathbb{P}[\overline{\mathbf{\xi}_{N(i)}}|\beta, \rho]} \\
&= \mathbb{P}[\xi_i = g|X_i, \beta] \frac{\mathbb{P}[\xi_i = g|\overline{\mathbf{\xi}_{N(i)}}, \rho]}{\mathbb{P}[\xi_i = g]} \\
&= \mathbb{P}[\xi_i = g|X_i, \beta] \frac{\frac{1}{Z_i^A} \exp[V_{ext}(z_i) + \rho_1 V_{loc}(\xi_i = g, \overline{\mathbf{\xi}_{N_1(i),g}}) + \rho_2 V_{reg}(\xi_i = g, \overline{\mathbf{\xi}})]}{\frac{1}{Z_i^B} \exp[V_{ext}(z_i)]} \\
&= \mathbb{P}[\xi_i = g|X_i, \beta] \frac{1}{Z_i} \exp[\rho_1 V_{loc}(\xi_i = g, \overline{\mathbf{\xi}_{N_1(i),g}}) + \rho_2 V_{reg}(\xi_i = g, \overline{\mathbf{\xi}_{N(i)}})] \\
&= \mathbb{P}[\xi_i = g|X_i, \beta] \mathbb{P}[\xi_i = g|\overline{\mathbf{\xi}_{N(i)}}, \rho, V_{ext} = 0]
\end{aligned}$$

Thus mean-field approximation of the likelihood can be written:

$$\mathbb{P}^{MF}[\mathbf{Y}|\mathbf{X}, \Theta, \beta, \rho] = \prod_{i=1}^n \sum_{g=1}^G \mathbb{P}[Y_i|\xi_i = g, \theta_g] \mathbb{P}[\xi_i = g|X_i, \beta] \mathbb{P}[\xi_i = g|\overline{\mathbf{\xi}_{N(i)}}, \rho, V_{ext} = 0]$$

C. IMPLEMENTATION

C.1 *Software*

The implementation was done with  software (R Development Core Team, Vienna, Austria, <http://www.r-project.org/>), using the `glm` function from package *stats* for the regression model, the `nnet` function of package *e1071* for the concomitant model. To perform optimization of the regularization model, we used a quasi-Newton algorithm (L-BFGS-B Limited memory Broyden Fletcher Goldfarb Shanno with box constraints) of the `optim` function from the *stats* package. To speed up the computations, the spatial regularization part of the E-step was implemented in C++ and linked to  using the *Rcpp* package.

C.2 *Spatial regularization algorithm*

The estimation of the spatial regularization of the posterior probabilities was based on the Iterated Conditional Mode algorithm (Besag (1986)) adapted for mean-field approximation (Celeux *and others* (2003)). At each step, a site ordering was randomly selected. Then, for each site in turn $\mathbb{P}[\xi_i = g | \overline{\xi_{N(i)}}, \rho]$ was computed for each group using equation 3.4 and the corresponding posterior probability updated. A step ended when all the sites have been visited once. Convergence for local regularization was considered achieved when the maximum of the absolute differences of the posteriors between two consecutive estimations over all sites was low. The threshold value was chosen at 0.01.

Because the calculation of the regional potential is computationally expensive, we updated this potential only at the end of a step when the posteriors have converged. The spatial regularization step was stopped at stabilization of both the regional potential and the posterior probabilities. Note that the number of neighbors increases quickly with the increase in distance. For example, with a regular 3D lattice and unit distances between sites, a given site has more than 4000

neighbors within a range of 10 voxels. In practice, an upper limit should be fixed for the maximum-order neighborhood C_{max} to enable the computation in case of large spatial groups.

D. EVALUATION ON 3D ARTIFICIAL DATA

D.1 *Data generation and processing*

The simulation scenario presented in section 5.1 was adapted for 3D data. Eleven slices were considered which led to a regular lattice of $90 \times 90 \times 11$ sites. The same three groups presented in section 5.1 were used. The artifacts of Dataset 2 and 3 were generated in 3D with decreasing intensity with the increase of distance from the central slice (see figure B and figure C). The assessment of the model performance was based on Brier score and the simple matching coefficient.

The results are summarized in table A and are consistent with those found with the 2D simulation. Figure D compares the posterior probabilities of Group 3 with various specifications of the finite mixture model on three slices of Dataset 3. As in 2D, unlike the regional regularization, the local regularization could not remove large-scale artifacts.

Dataset	Performances	$\rho=(0,0)$	$\rho=(6,0)$	$\rho=(12,0)$	$\rho=(0,6)$	$\rho=(6,6)$
		Brier score (%)				
Dataset 1	Global	22.8	3.45	1.29	6.69	2.32
Dataset 2	Global	23.9	7.32	5.32	8.19	4.69
Dataset 2	Artifact sites	75.6	64.7	51.5	48.4	40.9
Dataset 3	Global	24.3	9.59	6.47	6.83	2.32
Dataset 3	Artifact sites	70.3	67.4	47.1	15.5	0.519

Dataset	Performances	Simple matching coefficient				
Dataset 1	Global	93.2	99.9	100.0	99.4	99.9
Dataset 2	Global	92.6	99.4	99.7	99.1	99.8
Dataset 2	Artifact sites	29.3	54.9	72.7	72.7	82.2
Dataset 3	Global	92.4	99.0	99.6	99.4	99.9
Dataset 3	Artifact sites	39.6	50.6	76.8	96.8	100.0

Table A. Evaluation of Group 3 classification on the 3D simulated data with the finite mixture model (fMM) and different specifications of the spatial regularization parameter ρ . Dataset 1 was used to test the robustness to heterogeneity in group intensity. Dataset 2 was used to test the robustness to linear artifacts. Dataset 3 was used to test the robustness to spherical artifacts.

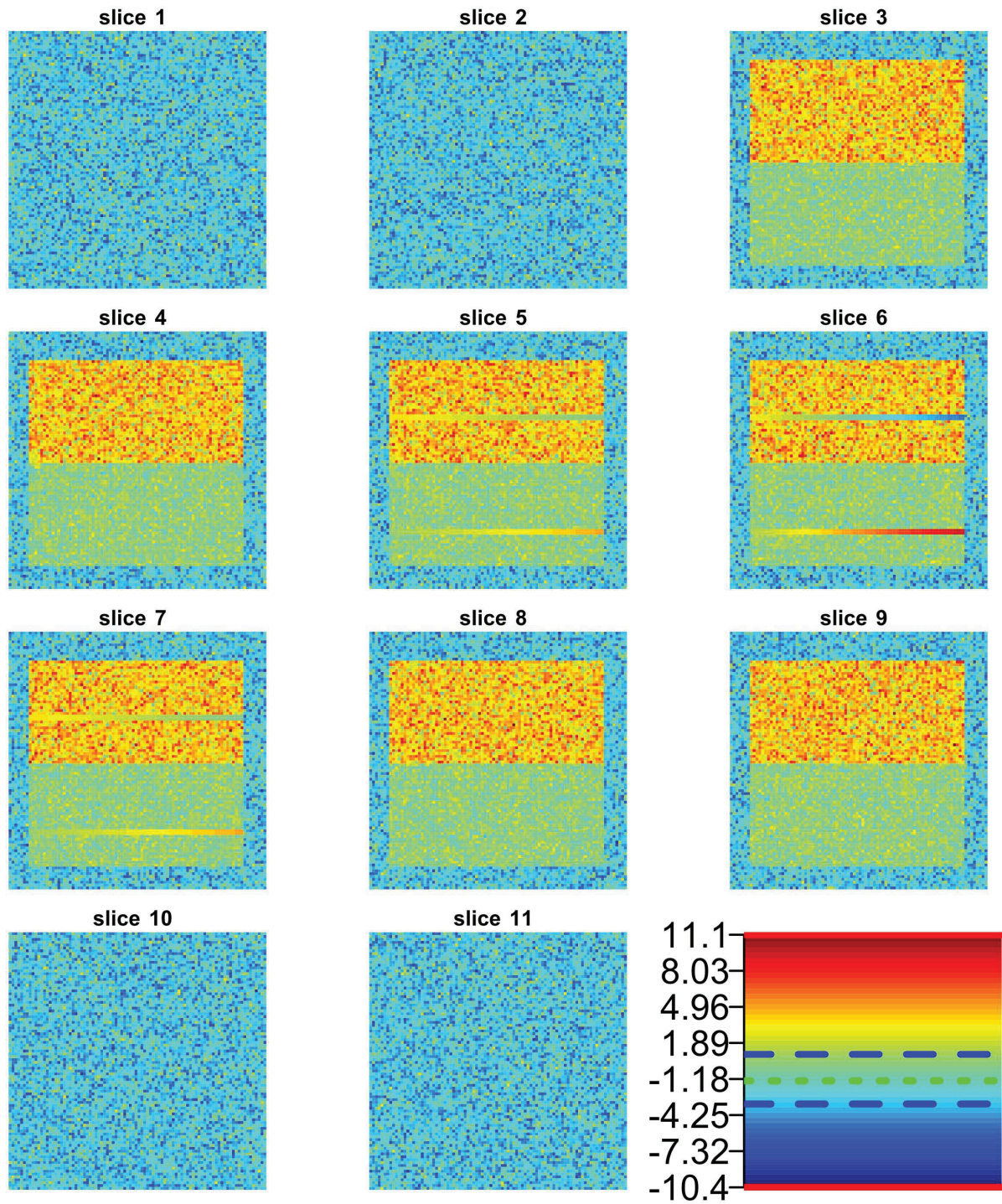


Fig. B. Three-dimensional Dataset 2. Each image corresponds to a slice of the dataset (from slice 1 to slice 11).

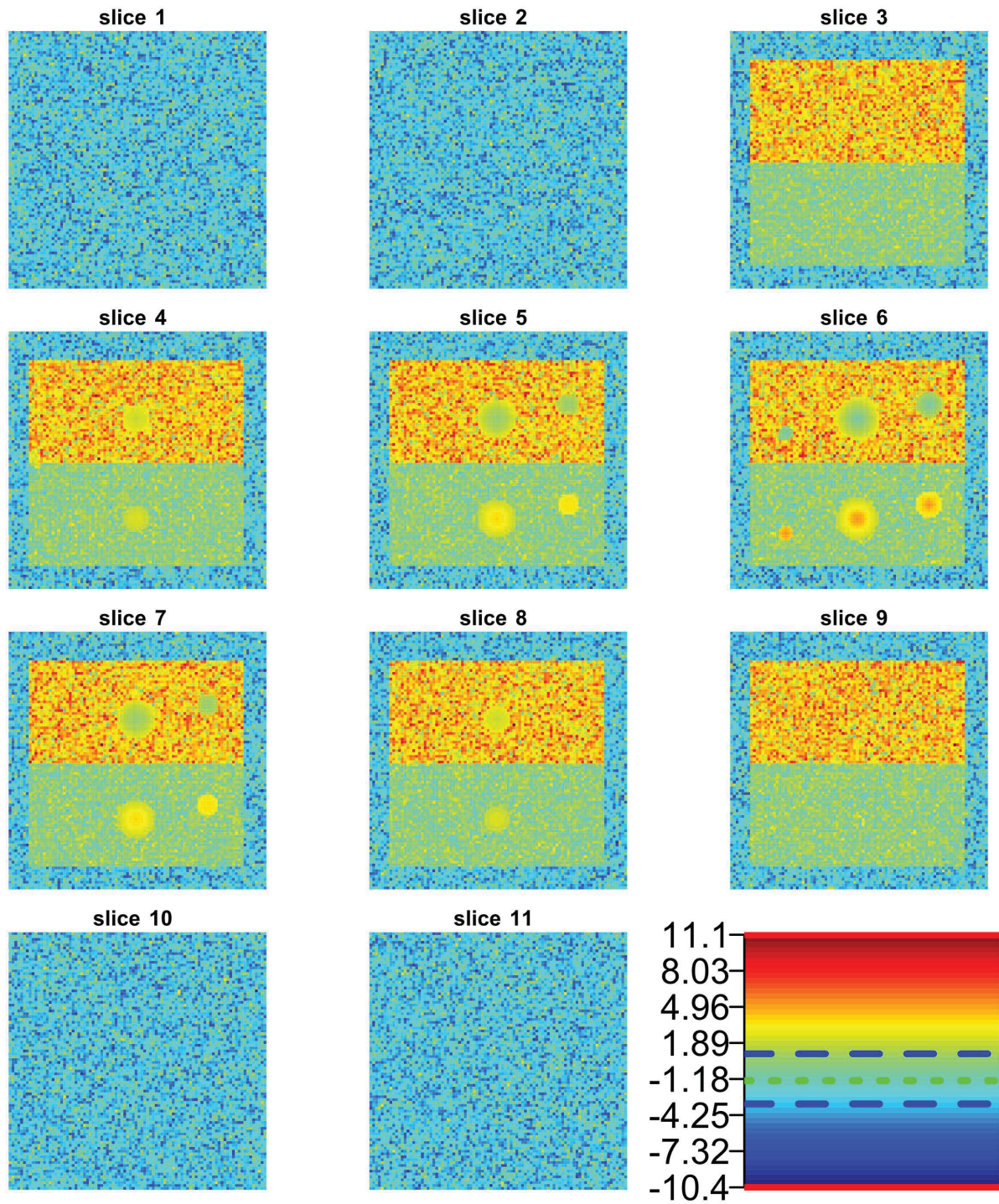


Fig. C. Three-dimensional Dataset 3. Each image corresponds to a slice of the dataset (from slice 1 to slice 11).

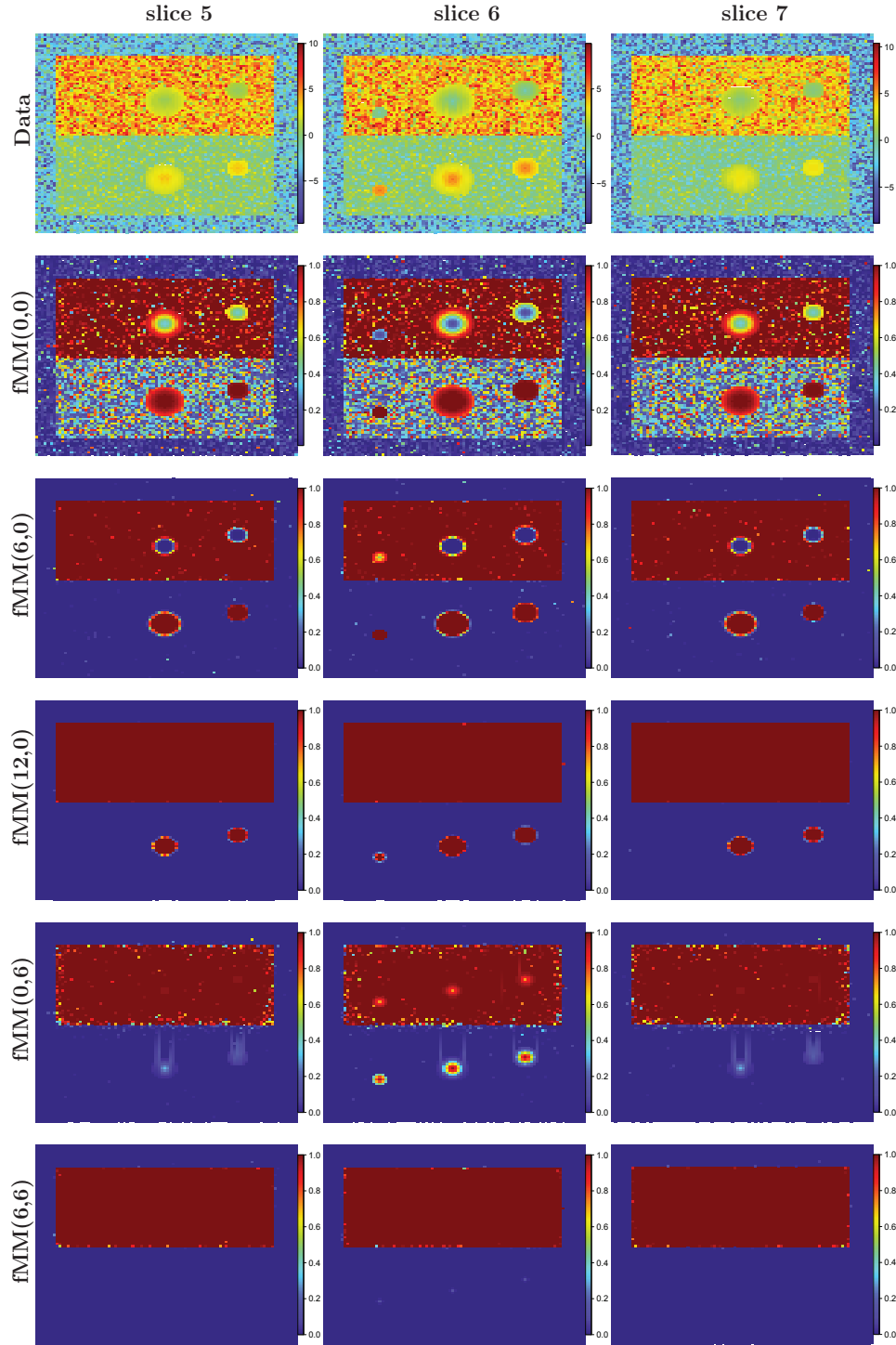


Fig. D. Posterior probabilities of Group 3 membership for the five specifications of the fMM and the three simulated datasets displayed for three slices of the 3-dimensional Dataset 3.

E. BIAS OF THE REGULARIZATION PARAMETER ESTIMATOR

E.1 *Simulation study*

The bias of the regularization parameter estimator was investigated using a three class Potts model restricted to immediate dependencies (see Section 3.1). The simulation of the Potts model was performed using a Gibbs sampler, as proposed in Gaetan and Guyon (2000), with 1000 iterations. Example of the simulated fields are represented on figure E. The number of observations was varied from 100 to 30625 and the local regularization parameter ρ_1 was varied from 0 to 10. Each scenario was replicated 250 times.

The boxplots of the relative bias for three sample sizes across the local regularization parameter values are represented on figure F. As expected from the theory, the variance and the bias of the estimator decreased with the sample size. For a sample size of 30625 the maximum relative bias was 0.001 reached for $\rho = 10$. The root mean square errors ranged from 0.002 to 0.02. Then for 99.9% of the simulations the relative difference between the true parameter value and the estimated parameter value was below 5%. As typical MRI images have large number of voxels (in our case in median 30439 voxels [minimum 22657 ; maximum 36181]), the estimations should merely be slightly biased.

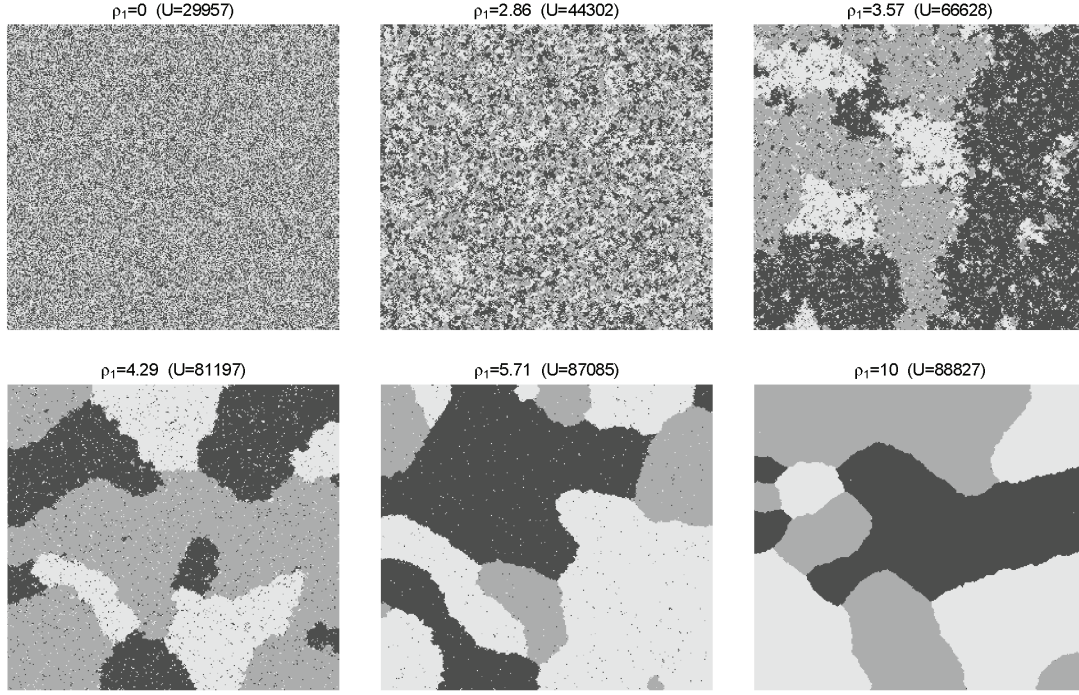


Fig. E. From left to right, top to bottom : an example of the simulated field using a Gibbs sampler after 1000 iterations for various values of the Potts parameter (ρ_1) and the corresponding potential (U).

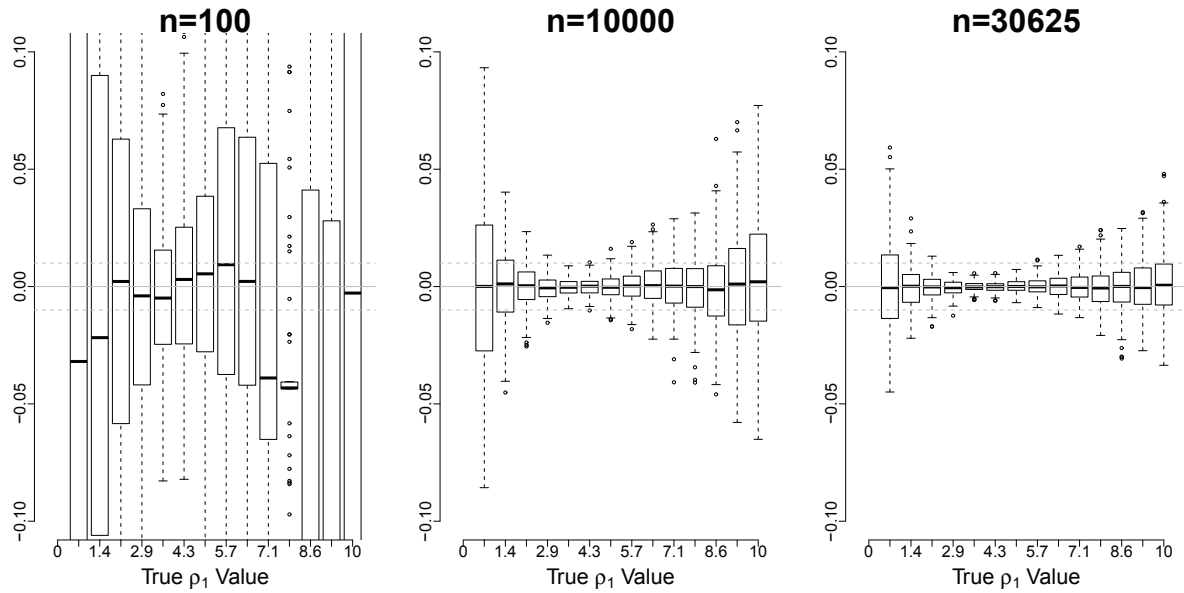


Fig. F. From left to right : boxplot of the relative difference between the true and the estimated local regularization parameter for $n=100$, 10000 and 30625. The grey solid line indicate no bias and the dashed solid lines a relative bias of $\pm 1\%$

E.2 *Comparison between internal and external estimation of the regularization parameters*

In the article, we proposed two methods for estimating the spatial parameters : internal estimation and external estimation (see paragraph 2 of section 6.2). To study their validity and retain the more appropriated one, we compared the parameters estimated by each method with those estimated on the physicians' masks.

The differences are displayed on figure G. Both methods tended to over-estimate the local parameter and under-estimate of the regional parameter. However, the difference was much greater with the internal than with the external method. The median difference for the local regularization parameter of fMM_{reg1} was 2.12 (interquartile range : 0.68) with the internal estimation and 0.17 (0.07) with the external estimation. The differences were similar when considering the local regularization parameter of fMM_{reg2} : 2.24 (0.76) vs. 0.21 (0.05). For the regional regularization parameter of fMM_{reg2} , the difference was -1.33 (0.78) for the internal estimation and -0.30 (1.03) for the external estimation. This difference was much greater than the estimator bias: an imperfect knowledge of the group membership appeared to be the predominant cause of the inaccuracy of the estimator compared to the bias introduced by mean-field approximation.

The proposed segmentation algorithm was rather robust to the specification of the regularization parameters. Overall, the median relative difference in the estimated volumes with the internal estimation compared to external estimation was 0.52% with an inter-quartile range of 3.0%. However, there were large discrepancies between the estimated volumes of the two methods in two patients (Patient 9 and 12). In both cases, a much lower regional regularization parameter was estimated by the internal estimation vs. the external estimation (0 and 1.18 compared to 2.61) that limited artifact correction.

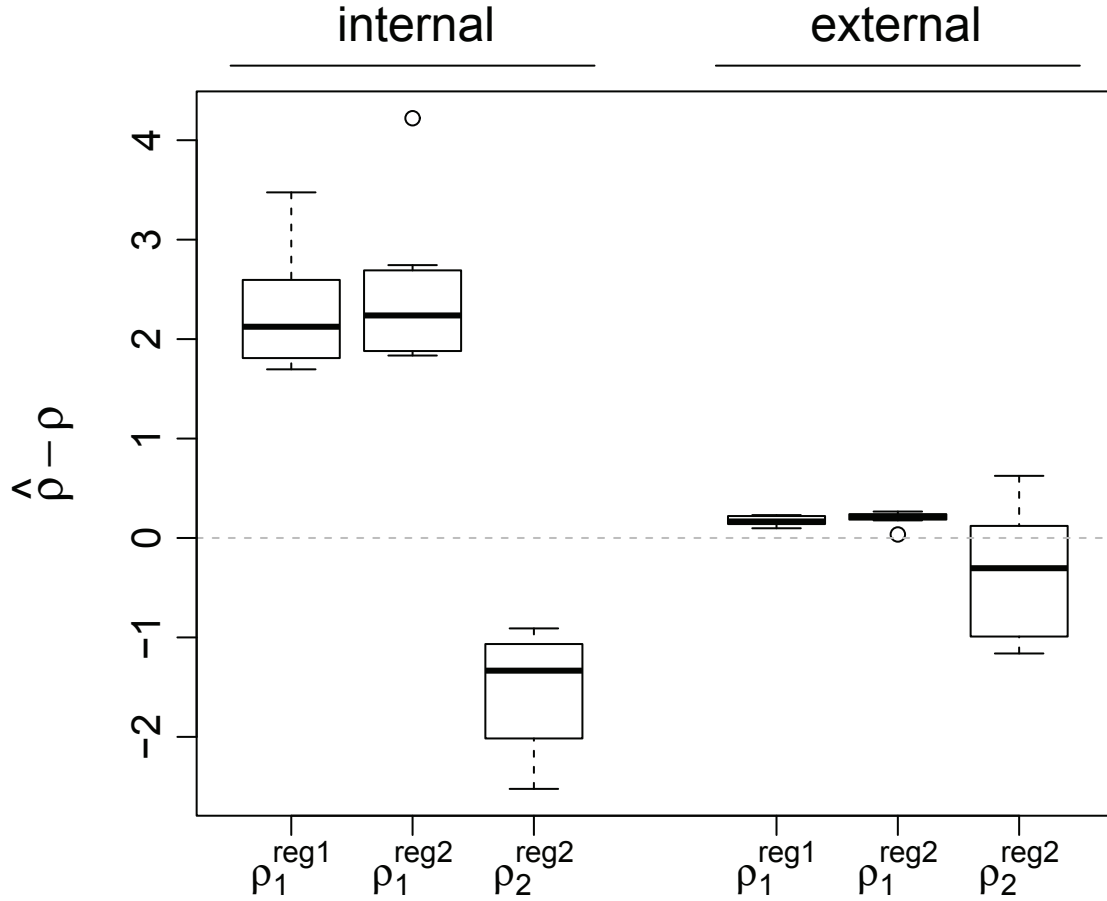


Fig. G. Difference between the regularization parameters estimated on the finite mixture model segmentation ($\hat{\rho}$) and those estimated on the physicians' masks (ρ). ρ_1^{reg1} corresponds to the local regularization parameter of the finite mixture model with local regularization. ρ_1^{reg2} and ρ_2^{reg2} correspond respectively to the local and the regional regularization parameter of the finite mixture model with local and regional regularization. The first three boxplots correspond to the internal estimation method and the next three to the external estimation method.

F. POTENTIAL EVALUATION FOR FRAGMENTED LESIONS

In case of multiple lesion, computation of the potential is performed as follows:

- **Identification of the spatial groups** : pixels with membership to the lesion class above a pre-specified threshold (e.g. 0.1) are considered as lesioned. The lesion is then divided into spatial groups (i.e. groups without common neighbors).
- **Computation of the potentials** : for each spatial group that contains the considered pixel in its range (i.e. the pixel belongs to the neighborhood of at least one pixel of the group), the regional potential is computed.
- **Computation of the final potential** : for each pixel the highest potential is retained.

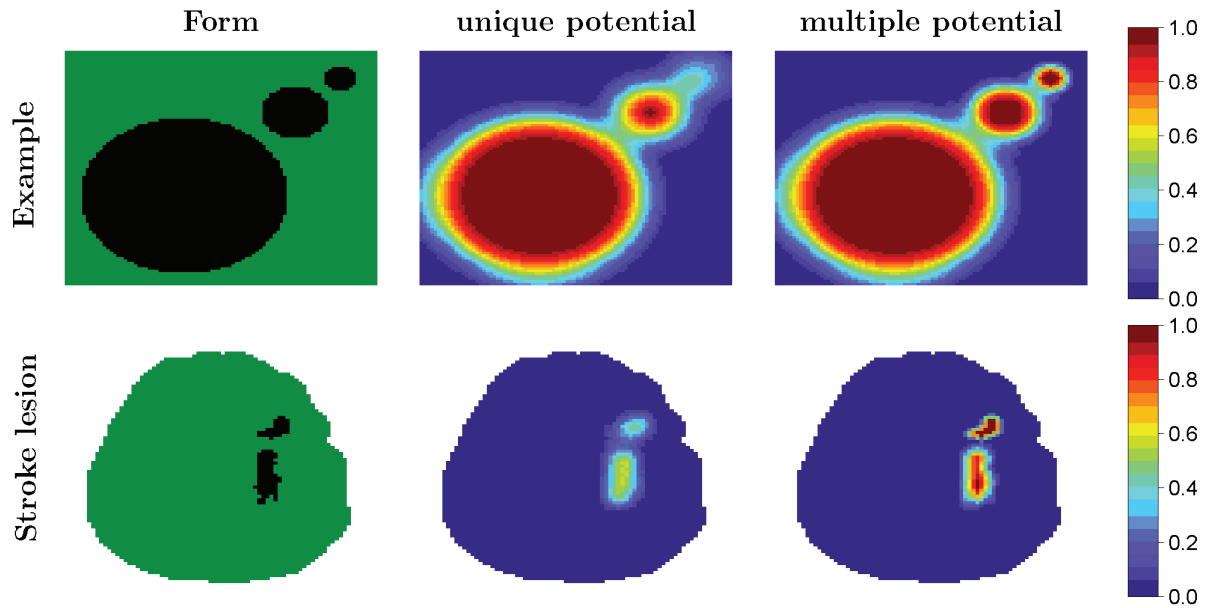


Fig. H. From left to right : form, regional potential with only one spatial group considered and regional potential adapted for multiple spatial groups, as described above. The first line is an example of geometry and the second line an example of 3D fragmented stroke lesion, for which one slice is represented.

G. REPRESENTATION OF PATIENT 5, 11 AND 12

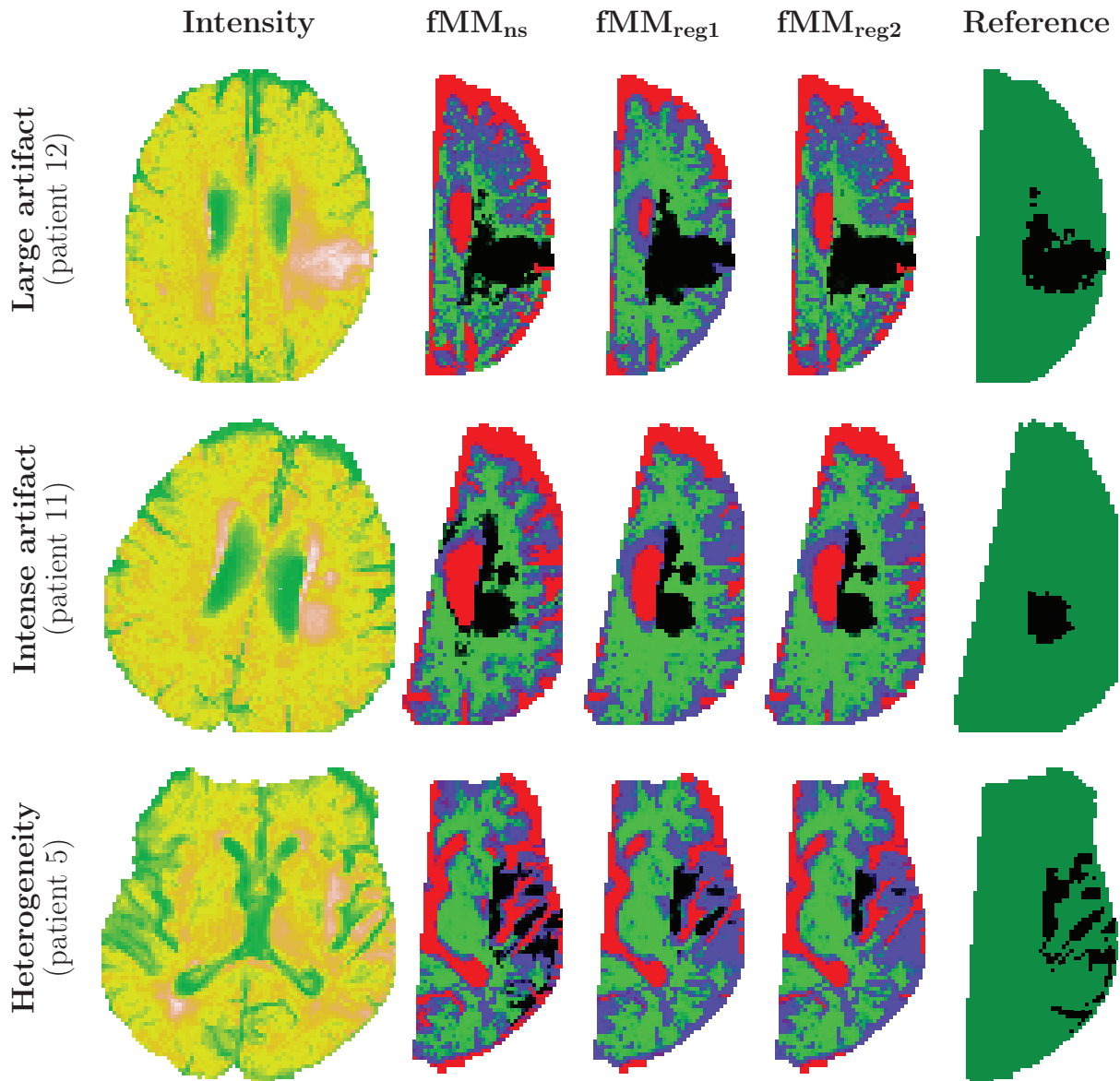


Fig. I. Results of the mixture models applied to Patient 5, 11 and 12. From left to right: T2 FLAIR cartography, estimated membership probability for the three mixture models (CSF in red, WM in blue, GM in green, lesion in black) and the corresponding physicians' mask (lesioned tissue in black, healthy tissue in green).

H. COMPARISON BETWEEN THE FMM WITH LOCAL REGULARIZATION AND THE FAST SOFTWARE

The fmm with only local regularization (presented here) and FAST software rely on the same model (see Zhang *and others* (2001) for a description of FAST algorithm), the only difference is that the concomitant model and the growing region step are not available in FAST. To compare the two algorithms we used the following options for FAST software:

- Four tissue-type classes
- Two image channels (one for T2 FLAIR intensity and one for T2 FLAIR contro intensity)
- MRF β value varying from 0.05 to 2 (the default FAST value is 0.1).
- Partial volume classification

In comparison with our algorithm, FAST software was faster; the computation time was about 1 minute for each patient. In term of model performance, FAST had a lower Jaccard coefficient vs. the segmentation with our locally regularized fmm (see figure J). This was true for all MRF β parameters tested, despite the improvement seen when β was increased from 0.05 to 2. The optimal β was patient-dependent (e.g., $\beta=1.5$ for Patient 5, $\beta=0.5$ for Patient 8 and $\beta=2$ for Patient 7). The main reason for the lower performance of FAST was its sensitivity to artifacts caused by high-intensity non-lesioned cerebral structures. Such artifacts are mostly present in the lower slices and were removed by the growing region step of our algorithm. Nevertheless, similar segmentations were obtained when we compared FAST segmentation with fmm using only local regularization on central slices (see figure K).

The difference in computation time between the two algorithms can be explained by the fact that FAST software has been specifically designed for modeling Gaussian intensity distributions whereas our algorithm allows a broader range of distributions with adjustment on covariates, a concomitant model, and a growing region step. Finally, the code associated with FAST has been

more optimized than ours.

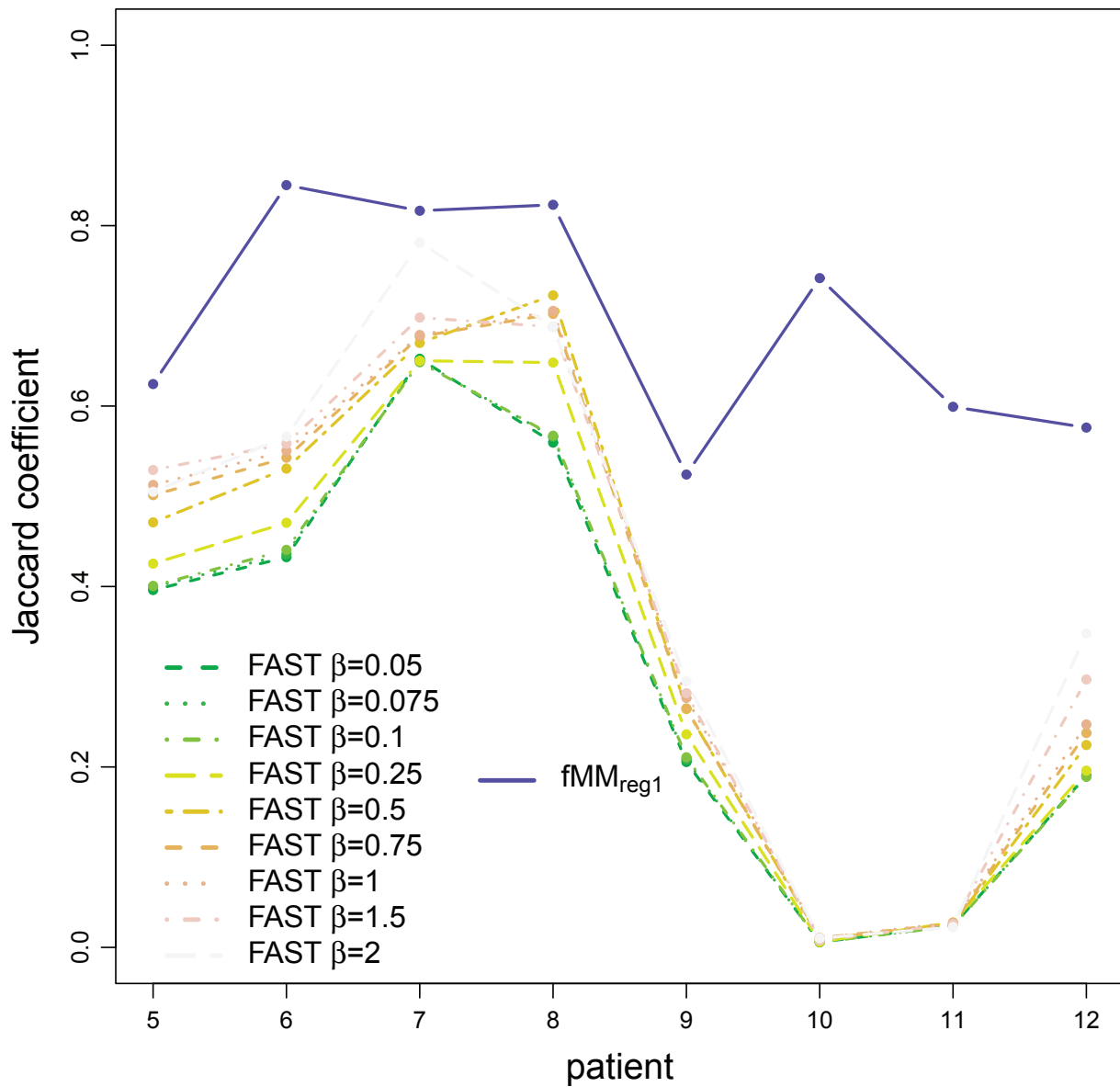


Fig. J. Quality of the segmentation evaluated by Jaccard coefficient for Patients 5 to 12 using the fmm with local regularization (blue line) and various specifications of FAST software (green to white line). FAST software was used varying the spatial parameter β from 0.05 to 2.

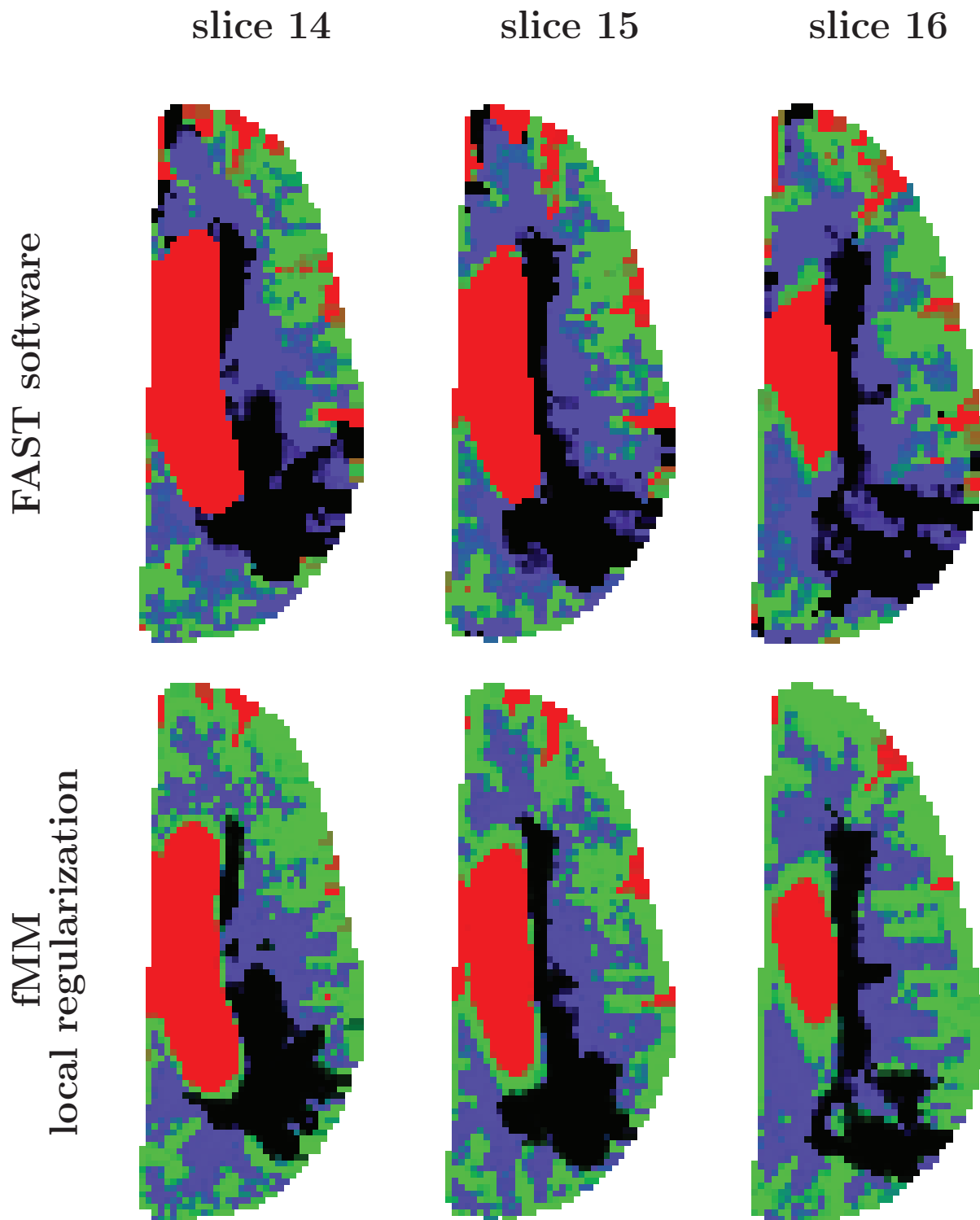


Fig. K. Comparison between the segmentation of FAST software (with $\beta = 1$, first row) and the segmentation of the fMM with local regularization (second row) on three slices of Patient 9. Voxels classified as lesioned are displayed in black.

REFERENCES

- BESAG, J. (1986). On the statistical analysis of dirty pictures. *Journal of the royal statistical society. Serie B (Methodological)* **48**(1), 259–302.
- CELEUX, G., FORBES, F. AND PEYRARD, N. (2003). EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pattern Recognition* **36**(1), 131 – 144.
- CHANDLER, D. (1987). *Introduction to Modern Statistical Mechanics. Chapter 5*. Oxford University Press.
- DEMPSTER, A. P., LAIRD, N. M. AND RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Serie B (Methodological)* **39**(1), 1–38.
- GAETAN, C. AND GUYON, X. (2000). *Spatial Statistics and Modeling*. Springer.
- GUNAWARDANA, A. AND BYRNE, W. (2005). Convergence theorems for generalized alternating minimization procedures. *Journal of Machine Learning Research* **6**, 2049–2073.
- HOFMANN, T. AND BUHMANN, J.M. (1997). Pairwise data clustering by deterministic annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(1), 1–14.
- KABIR, Y., DOJAT, M., SCHERRER, B., FORBES, F. AND GARBAY, C. (2007). Multimodal MRI segmentation of ischemic stroke lesions. *29th Annual International Conference of the IEEE* **1**, 1595–1598.
- ZHANG, J. (1992). The mean-field theory in EM procedures for Markov random fields. *IEEE Transactions on Signal Processing* **40**(10), 2570–2583.

ZHANG, Y., BRADY, M. AND SMITH, S. (2001). Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging* **20**(1), 45–57.

Comparaison par rapport aux méthodes récemment publiées

Alors qu'au début de notre travail aucune méthode de segmentation ne tenait compte des difficultés liées à la maladie de la substance blanche, trois travaux ont été publiés durant la phase de publication de notre méthode. N'étant pas évoqués dans l'article, nous les décrivons ici brièvement et discutons de leur intérêt.

Le premier travail (DALCA et al., 2014), propose une approche par modèle de mélange mais avec des différences par rapport au modèle que nous proposons. Tout d'abord, il restreint l'espace d'étude à la substance blanche en éliminant les voxels du liquide cérébro-spinal et de la substance grise.

Cela permet de ne considérer que trois groupes : le tissu sain, la lésion et la maladie de la substance blanche (leucoaraiose dans l'article). Leur modèle se compose de trois sous-modèles :

- un modèle d'intensité Gaussien univarié utilisant le T2 FLAIR.
- un modèle spatial correspondant au modèle de Potts standard (régularisation locale uniquement). Le paramètre spatial du modèle est fixé a priori.
- un a priori portant sur la distribution spatiale qui identifie les voxels correspondant à la maladie de la substance blanche. Pour cela, une analyse en composantes principales est réalisée sur des images tests pour identifier les directions caractéristiques de la maladie de la substance blanche. Puis les probabilités a posteriori d'appartenir à la lésion sont projetées sur ces composantes afin d'éliminer les voxels correspondant à la maladie de la substance blanche.

L'algorithme de segmentation a été validé sur 6 volumes et 100 images. Cette approche propose une démarche originale dont certaines idées pourraient améliorer notre algorithme. Le point fondamental est l'utilisation d'un groupe additionnel "maladie de la substance blanche" qui doit permettre une meilleure caractérisation de cette maladie et une meilleure séparation avec le groupe lésion. En effet, en omettant ce groupe, nous obtenons des groupes de tissu sain très hétérogènes, ce qui nuit à la bonne identification de chacun des groupes. Notre modèle avait certes été testé avec 5 groupes mais sans résultats convaincants : l'algorithme a en effet besoin d'une variable discriminante pour le cinquième groupe. Pour cela, un a priori sur la distribution spatiale de la maladie de la substance blanche est utilisé dans l'article de DALCA et al., 2014. L'idée est intéressante mais contraignante : il faut normaliser le cerveau vis-à-vis d'un cerveau de référence, ce qui peut être source d'imprécisions. De plus, la géométrie de la maladie de la substance blanche varie d'un patient à l'autre (voir figure III.15), il peut donc être délicat de définir une géométrie universelle.

Le second travail (MITRA et al., 2014) est assez différent puisqu'il propose une approche supervisée fondée sur les forêts d'arbres décisionnels. Les arbres sont entraînés à distinguer la lésion, le tissu sain, et la maladie de la substance blanche en utilisant les paramètres IRM classiquement disponibles (T1, T2, T2 FLAIR et ADC) ainsi que des probabilités d'appartenance à la lésion, à la substance blanche et à la substance grise.

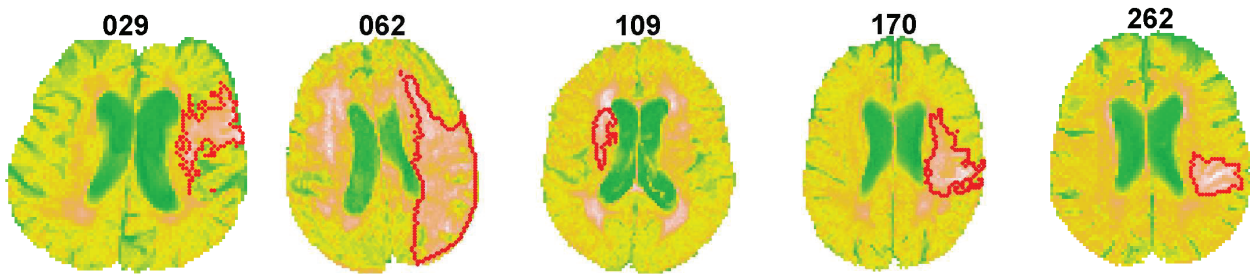


FIGURE III.15 – Cartographie T2 FLAIR de 5 patients ayant une maladie de la substance blanche marquée. La lésion identifiée par le médecin est contourée en rouge.

Ces probabilités sont déterminées à l'aide d'une séquence de deux modèles de mélange. La méthode est validée sur 36 patients avec des résultats modestes (erreur volumique relative de $32,32 \% \pm 21,64 \%$) mais qui restent intéressants au vu de la difficulté de la segmentation. Toutefois, comme le précisent les auteurs, la robustesse de la méthode face à la maladie de la substance blanche admet ses limites : "Age related WMH [hyperintensités causées par la maladie de la substance blanche] that appear around the periventricular regions exhibit similar hyperintensities as ischemic lesions on T2/FLAIR, and therefore are indistinguishable from stroke using our method.". En effet, dans leur modèle la distinction lésion et maladie de la substance blanche est quasi-exclusivement fondée sur un critère d'intensité. Or bien souvent les deux ont des intensités similaires.

Enfin le dernier travail (MAIER et al., 2015) est une approche de classification supervisée basée sur l'algorithme Extra-Tree. Il s'agit d'une méthode agrégeant plusieurs arbres de décisions afin d'améliorer leur capacité prédictives. Les arbres utilisés ne sont pas des arbres de décision classiques puisqu'en chaque nœud, le point de coupure de la variable est choisi de façon aléatoire (i.e. la valeur τ permettant de découper les données en deux groupes : $X > \tau$ et $X \leq \tau$). Seul le choix de la variable est fait de façon supervisée. En plus des variables d'intensité, plusieurs variables complémentaires sont utilisées pour améliorer la classification (intensité moyenne locale, distance au centre du cerveau). Cette méthode a été validée sur 37 patients et semble robuste aux artéfacts causés par la maladie de la substance blanche (voir figure 9 de l'article).

Limites de l'approche et perspectives

Comme évoqué dans la discussion de l'article, l'algorithme de segmentation que nous proposons admet un certain nombre de limites. Du point de vue de l'estimation, la méthode proposée n'est pas complètement satisfaisante :

- **sensibilité à l'initialisation** : nous utilisons un algorithme de type EM pour lequel il n'y a aucune garantie de converger vers l'optimum global. En pratique, suivant les initialisations, des solutions différentes peuvent être obtenues dont certains seront sous-optimales.

Une solution possible serait d'utiliser des méthodes de Monte-Carlo pour estimer le modèle (WOOLRICH et al., 2005; ZHANG et al., 2008; FORBES et al., 2007). Pour des raisons de temps de calcul n'avons pas considéré ce type d'approche dans cette thèse. Toutefois, de futurs travaux devraient s'intéresser à évaluer le gain (en terme de qualité de classification) et le coût (en terme de temps de calcul) d'une approche Monte-Carlo. Une troisième approche consisterait à exploiter l'information disponible dans l'hémisphère controlatéral. L'AVC étant généralement unihémisphérique, cet hémisphère peut être utilisé pour initialiser (ou fixer) les valeurs des paramètres associés aux tissus sains. En réduisant l'espace de recherche on peut espérer limiter le risque de converger vers un minimum non global.

- **l'approximation du champ moyen** : cette approximation modifie la fonction objectif et peut conduire à des solutions sous-optimales.
- **les écarts au cadre théorique** : pour des raisons pratiques nous avons introduit deux écarts au modèle proposé. Le premier est l'estimation automatique de C_{max} pour le calcul du potentiel régional. De ce fait, la forme du potentiel dépend de la taille du groupe. Le potentiel n'est alors plus seulement une fonction de cliques d'ordre au plus 2. Il n'y a plus de garantie que la constante optimale au sens de la divergence de Kullback-Leibler soit l'espérance. L'autre écart est l'utilisation d'un algorithme de croissance de régions pour éliminer les artéfacts présents sur les coupes basses. Toutefois, cette étape pourrait être remplacée par une pénalisation de l'appartenance à la lésion en fonction de la distance aux points de référence. Une telle pénalisation s'intégrerait mieux dans le modèle de mélange.
- **la résolution du problème de point fixe** : la méthode que nous employons pour résoudre ce problème n'est pas optimale. Ainsi CELEUX et al., 2003, montrent que l'algorithme en champ simulé donne de meilleures solutions. Cependant le gain reporté dans l'article nous semblait faible par rapport au coût numérique. STORATH et al., 2014, ont récemment proposé méthode de résolution efficace du problème de Potts, à paramètre spatial fixé. Aussi il nous a semblé intéressant de détailler dans la section suivante comment cette approche pourrait s'intégrer dans notre algorithme d'estimation. A noter que la résolution n'est toutefois proposée que pour le modèle de Potts avec potentiel local.

D'un point de vue pratique notre méthode nécessite de :

- **préciser un point d'initialisation** : soit en initialisant les paramètres de distribution de chaque groupe, soit en initialisant l'appartenance de chaque voxel aux groupes. Ce point est en fait facilement automatisable : dans l'article nous utilisons le résultat d'un algorithme des k-moyennes pour l'initialisation. Comme mentionné précédemment il pourrait être plus judicieux d'initialiser les paramètres de distribution de la substance grise, de la substance blanche, du liquide cérébro-spinal à partir de valeurs obtenues dans l'hémisphère controlatérale.

- **initialiser les zones de lésion** : un point par groupe spatial de lésion doit être précisé lorsque l'algorithme de croissance de lésion est utilisé. Ce point est nécessaire lorsque l'on utilise un paramètre d'intensité peu spécifique, comme le T2 FLAIR. Il pourrait néanmoins être automatisé en utilisant d'autres séquences, moins sensibles mais plus spécifiques, comme la séquence en diffusion.
- **fixer a priori le paramètre de régularisation régional** : en présence d'artéfacts, nous avons constaté que la procédure d'estimation automatique était biaisée. Fournir une valeur estimée sur des images sans artéfact est apparue plus satisfaisant.

Enfin, en effectuant une légère modification à l'algorithme proposé, il est possible d'intégrer un a priori spatial comme version simplifiée de ce que proposent DALCA et al., 2014. L'utilisateur préciserait une carte probabiliste de la présence de maladie de la substance blanche qui serait utilisée pour compléter l'a priori externe du modèle. Les voxels situés dans des régions où la maladie de la substance blanche est fréquente seraient alors moins facilement associés à la lésion.

1.7 Vers une meilleure estimation du modèle de Potts

L'estimation du modèle spatial est un problème récurrent dans les algorithmes de segmentation d'image. Il se pose à la fois lorsque l'on considère un modèle de mélange mais aussi pour les approches de déformation de formes. En effet, même si ces deux approches correspondent à des points de vue différents, leur fonction objectif présente des similitudes : dans les deux cas, il s'agit d'optimiser un compromis entre un terme d'adéquation aux données et un terme de régularité. Une méthode de résolution globale a été récemment proposée pour le problème de Potts (STORATH et al., 2014), c'est-à-dire que cette méthode possède des garanties supérieures de converger vers un optimum global comparée à notre méthode de résolution. Nous allons donc chercher à lier notre fonction objectif à celle utilisée dans l'article de STORATH et al., 2014, en identifiant les hypothèses effectuées.

Redéfinition de la fonction objectif

Comme précédemment nous allons faire l'hypothèse que la distribution spatiale du modèle est factorisable. L'approximation du champ moyen conduit à écrire la distribution approchée suivante :

$$\mathbb{P}^{MF} [\mathbf{Y}|\boldsymbol{\Theta}] = \prod_{i=1}^n \sum_{g=1}^G \mathbb{P} [Y_i|\xi_i = g, \theta_g] \mathbb{P} [\xi_i = g|\overline{\boldsymbol{\xi}_{\mathbf{N}(i)}}]$$

Supposons des distributions gaussiennes pour le modèle d'intensité :

$$\begin{aligned} \mathbb{P}^{MF} [\mathbf{Y}|\boldsymbol{\Theta}] = \prod_{i=1}^n \sum_{g=1}^G \frac{1}{\sqrt{2\pi\sigma_g^2}} e^{-\frac{(Y_i - \mu_g)^2}{2\sigma_g^2}} \frac{1}{Z_i} \exp \left(\rho_{local} V_{local}(\xi_i = g|\overline{\boldsymbol{\xi}_{\mathbf{N}_{local}(i)}}) \right. \\ \left. + \rho_{régional} V_{régional}(\xi_i = g|\overline{\boldsymbol{\xi}_{\mathbf{N}_{régional}(i)}}) \right) \end{aligned}$$

Supposons une variance égale entre les groupes et l'absence d'interactions régionales ($\rho_{\text{regional}} = 0$). Les Z_i étant dépendant des valeurs de ρ_{local} , nous noterons $Z^*(\rho_{\text{local}}) = \prod_{i=1}^n \frac{1}{Z_i}$. On obtient alors :

$$\mathbb{P}^{MF}[\mathbf{Y}|\Theta] = \frac{1}{Z^*(\rho_{\text{local}})\sqrt{2\pi\sigma^2}^n} \prod_{i=1}^n \sum_{g=1}^G \exp\left(-\frac{(Y_i - \mu_g)^2}{2\sigma^2} + \rho_{\text{local}} V_{\text{local}}(\xi_i = g | \overline{\xi_{\mathbf{N}_{\text{local}}(i)}})\right)$$

Algorithme EM classifiant

L'approche proposée dans l'article de STORATH et al., 2014, est une approche déterministe, c'est-à-dire que l'appartenance des voxels aux groupes n'est pas probabiliste mais binaire. Afin de se ramener à ce cadre nous allons considérer un algorithme EM classifiant (ou CEM) : il s'agit d'une modification de l'étape E où chaque observation est rattachée au groupe le plus probable. Même si l'algorithme ne bénéficie plus des propriétés de l'algorithme EM classique, il a des propriétés analogues pour un autre type de vraisemblance : la vraisemblance classifiante (voir l'article de CELEUX et al., 1991). Sous l'approximation du champ moyen, ce nouvel algorithme s'écrit :

étape E : trouver l'appartenance a posteriori des voxels aux groupes :

$$\hat{\xi} = \underset{\xi}{\operatorname{argmax}} \left(\mathbb{P}^{MF}[\xi | \mathbf{Y}, \hat{\mu}, \hat{\sigma}, \hat{\rho}] \right)$$

où $\hat{\mu}$, $\hat{\sigma}$, $\hat{\rho}$ sont les estimations obtenues à l'étape M des paramètres du modèle.

étape M : trouver les paramètres associés à la distribution de probabilité de chacun des groupes :

$$(\hat{\mu}, \hat{\sigma}, \hat{\rho}) = \underset{\mu, \sigma, \rho}{\operatorname{argmax}} \left(\sum_{i=1}^n \sum_{g=1}^G \widehat{\xi}_{i,g} \left(-\frac{(Y_i - \mu_g)^2}{2\sigma^2} + \rho_{\text{local}} V_{\text{local}}(\xi_i = g | \overline{\xi_{\mathbf{N}_{\text{local}}(i)}}) \right) - \log(Z^* \sqrt{2\pi\sigma^2}^n) \right)$$

où $\widehat{\xi}_{i,g}$ est l'estimation obtenue à l'étape E du groupe d'appartenance du site.

L'étape M étant analogue à celle de l'algorithme de mélange présenté dans l'article, nous allons nous focaliser sur l'étape E.

De l'étape E vers le modèle de Potts

Explicitons l'étape E :

$$\begin{aligned} \hat{\xi} &= \underset{\xi}{\operatorname{argmax}} \left(\mathbb{P}^{MF}[\xi | \mathbf{Y}, \mu^k, \sigma^k, \rho^k] \right) \\ &= \underset{\xi=(\xi_1 \dots \xi_n)}{\operatorname{argmax}} \left(\frac{1}{Z} \prod_{i=1}^n e^{-\frac{(Y_i - \mu_{\xi_i})^2}{2\sigma^2} + \rho_{\text{local}} V_{\text{local}}(\xi_i = \xi_i | \overline{\xi_{\mathbf{N}_{\text{local}}(i)}})} \right) \end{aligned}$$

Les paramètres du modèle de mélange étant fixés, Z est une constante. En appliquant la fonction logarithme à la fonction à maximiser on obtient :

$$\begin{aligned}
\hat{\xi} &= \operatorname{argmax}_{\xi=(\xi_1 \dots \xi_n)} \left(\sum_{i=1}^n -\frac{(Y_i - \mu_{\xi_i})^2}{2\sigma^2} + \rho_{local} V_{local}(\xi_i = g|\overline{\xi_{N_{local}(i)}}) \right) \\
&= \operatorname{argmin}_{\xi=(\xi_1 \dots \xi_n)} \left(\sum_{i=1}^n \frac{(Y_i - \mu_{\xi_i})^2}{2\sigma^2} - \rho_{local} V_{local}(\xi_i = g|\overline{\xi_{N_{local}(i)}}) \right) \\
&= \operatorname{argmin}_{\xi=(\xi_1 \dots \xi_n)} \left(\sum_{i=1}^n \frac{(Y_i - \mu_{\xi_i})^2}{2\sigma^2} + \rho_{local} (1 - V_{local}(\xi_i = g|\overline{\xi_{N_{local}(i)}})) \right) \\
&= \operatorname{argmin}_{\xi=(\xi_1 \dots \xi_n)} \left(\sum_{i=1}^n (Y_i - \mu_{\xi_i})^2 + \gamma (1 - V_{local}(\xi_i = g|\overline{\xi_{N_{local}(i)}})) \right)
\end{aligned}$$

où $\gamma = \rho_1 * 2\sigma^2$.

Par définition de V_{local} :

$$\begin{aligned}
1 - V_{local}(\xi_i = g|\overline{\xi_{N_{local}(i)}}) &= 1 - \sum_{j \in N_{local}(i)} w_{i,j}^{local} {}^t \xi_i \xi_j \\
&= \sum_{j \in N_{local}(i)} w_{i,j}^{local} - w_{i,j}^{local} {}^t \xi_i \xi_j \\
&= \sum_{j \in N_{local}(i)} w_{i,j}^{local} (1 - {}^t \xi_i \xi_j) \\
&= \frac{1}{\sum_{j \in N_{local}(i)} 1} \sum_{j \in N_{local}(i)} (1 - {}^t \xi_i \xi_j)
\end{aligned}$$

$1 - {}^t \xi_i \xi_j$ compte le nombre de fois que les sites i et j sont de groupe différents. $1 - V_{local}$ est donc proportionnel à $\|\nabla \xi_i\|_0$ la norme 0 du gradient de ξ_i . En considérant excluant les sites au bord du domaine (qui ont un nombre variable de voisins), le problème à résoudre se ré-écrit :

$$\hat{\xi} = \operatorname{argmin}_{\xi=(\xi_1 \dots \xi_n)} \left(\sum_{i=1}^n \gamma \|\nabla \xi_i\|_0 + \|\mu_{\xi_i} - Y_i\|_2^2 \right)$$

Cela revient à chercher une fonction u continue par morceaux vérifiant :

$$u^* = \operatorname{argmin}_u \left(\sum_{i=1}^n \gamma \|\nabla u_i\|_0 + \|u_i - Y_i\|_2^2 \right)$$

où u_i désigne la valeur prise par la fonction u au site i .

On retrouve donc le modèle de Potts résolu dans l'article de (STORATH et al., 2014, formule 1.1). Nous avons donc montré qu'il était possible d'utiliser l'algorithme proposé par STORATH et al., 2014, pour résoudre le modèle de Potts (à la place de l'algorithme ICM).

Néanmoins cela nous a conduit à faire un certain nombre d'hypothèses en plus de l'approximation du champ moyen :

- la distribution de l'intensité dans les groupes suit une loi normale (éventuellement multivariée) dont la variance est homogène entre les groupes.
- validité de la vraisemblance classifiante

Pour des raisons logicielles, nous n'avons pas pu pour l'instant intégrer cette méthode d'optimisation dans notre méthode de segmentation. Néanmoins, nous avons pu évaluer son intérêt sur les coupes de différents patients (voir figure [III.16](#)). Elle offre une cohérence spatiale qui, visuellement, semble meilleure que celle permise par l'approximation de champ moyen. Elle paraît particulièrement intéressante pour corriger l'hétérogénéité du signal en T2. Néanmoins, étant initialement développée pour du traitement photographique, elle n'est pour l'instant disponible qu'en 2D et pour des domaines rectangulaires. Son adaptation en 3D est en cours de développement par l'auteur de la méthode.

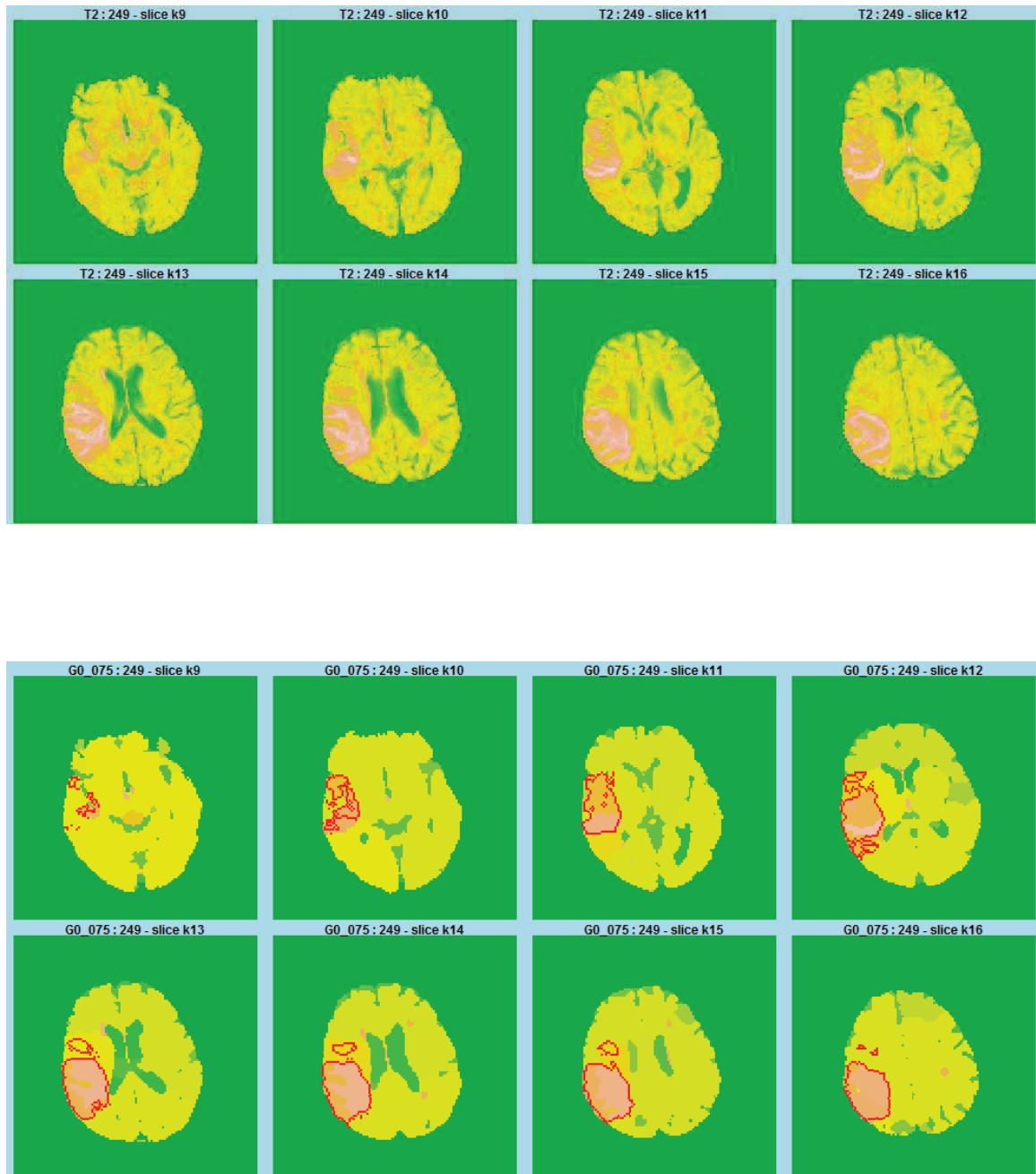


FIGURE III.16 – Exemple de résolution globale du modèle de Potts. En haut : 8 cartographies T2 FLAIR du patient appelé "7" dans l'article de segmentation. En bas : résolution du modèle de Potts avec $\gamma = 0,75$ sur ces 8 cartographies. Le contourage de la lésion par le médecin est affiché en rouge.

2 Développement logiciel pour l'AVC



2.1 Motivations



Les travaux effectués dans les sections précédentes ont nécessité des étapes de pré-traitements d'images pour disposer de données exploitables. Les pré-traitements permettent d'isoler les voxels correspondant au cerveau, de normaliser les données, de retraiter les paramètres IRM pour mieux extraire l'information qu'ils contiennent et de tenir compte des structures anatomiques dans l'analyse. Cette chaîne de pré-traitements, appelée **pipeline de pré-traitements**, peut rendre difficile la réutilisation des méthodes proposées pour de nouvelles études. Aussi, nous avons souhaité proposer une librairie de fonctions permettant de reproduire facilement l'ensemble des pré-traitements utilisés.

De nombreux logiciels existent pour gérer et traiter les données d'imagerie. Parmi les plus connus :


- FSL (The FMRIB Software Library, <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>)
- FreeSurfer (<http://surfer.nmr.mgh.harvard.edu/>)
- SPM (Statistical Parametric Mapping, <http://www.fil.ion.ucl.ac.uk/spm/doc/>)
- MIPAV (Medical Image Processing, Analysis, and Visualization, <http://mipav.cit.nih.gov/>)
- MATLAB, "Image Processing Toolbox" (<http://fr.mathworks.com/products/image/>)
- Python, "scikit-image" (<http://scikit-image.org/>)

Cependant, excepté MATLAB et Python, la plupart de ces logiciels sont des logiciels spécialisés. Ils incluent donc une bibliothèque d'outils développés et optimisés pour l'analyse de données d'imagerie. Toutefois, les outils d'analyse statistique proposés sont souvent limités et peu modulables.




Du point de vue statistique, l'un des logiciels de référence en termes de développement méthodologique est le logiciel . Il s'agit à la fois d'un logiciel et d'un langage de programmation. Il permet donc d'écrire ses propres librairies de fonctions et de les mettre à disposition de la communauté via des dépôts publics comme le CRAN (cran.r-project.org/) ou bioconductor (www.bioconductor.org/). Un nombre impressionnant de librairies pour l'analyse de données ont ainsi été développées (6819 au 1^{er} juillet 2015) ce qui en fait un logiciel de choix pour l'analyse de données. Des librairies comme *Rcpp* (EDDELBUETTEL et al., 2011) ou *snowfall* (KNAUS, 2013) permettent de facilement intégrer du code C++ ou de paralléliser les calculs, ce qui rend  compétitif vis-à-vis des autres logiciels en terme de puissance de calcul. Il possède aussi un certain nombre de librairies pour l'analyse de données d'imagerie dont une, *fslr* (MUSCHELLI, 2015), interface le logiciel FSL.

Notre travail étant principalement centré sur l'analyse statistique, nous nous sommes tournés vers le logiciel . Toutefois, les librairies disponibles nous ont paru peu adaptées à la gestion et au pré-traitement de nos données. Comme le montre une revue des librairies  en imagerie médicale (voir le tableau III.2), les formats de manipulation de données disponibles sont univariés c'est-à-dire qu'ils ne permettent de stocker qu'un seul paramètre (éventuellement mesuré de façon répétée dans le temps). De plus, les métadonnées¹ que l'on peut stocker avec les données d'imagerie sont limitées. Cela rendait la gestion de données complexe dans notre application :

- Nous devons gérer 11 paramètres IRM mesurés à 4 temps différents soit un total variant entre 17 et 31 paramètres IRM par patient, suivant les capacités d'acquisition des centres hospitaliers. De plus, les méthodes de filtrage appliquées aux paramètres IRM pour améliorer leur capacité prédictive (section 3.4) vont former 25 nouveaux paramètres par patient.
- Les données de chaque patient doivent être normalisées relativement à une zone saine de l'hémisphère controlatéral. Cela suppose d'associer à chaque paramètre la localisation des hémisphères et du tissu cérébral.

Nous avons donc choisi de développer une nouvelle librairie  pour faciliter la gestion et le pré-traitement de données d'imagerie multivariée. Cette librairie, appelée *MRIdaggr* pour agrégation de données IRM, est disponible sur le CRAN depuis mars 2015. Les deux sous-sections suivantes présentent le fonctionnement de la librairie ainsi que les fonctions utilisées pour effectuer le pipeline de pré-traitements.

2.2 Présentation de la librairie de fonctions MRIdaggr


La librairie MRIdaggr met à profit la programmation orientée objet pour rendre plus intuitif l'utilisation de données d'imagerie en langage  (voir le document de GENOLINI, 2009, pour une introduction à la programmation orientée objet sous ). Pour cela, elle implémente une nouvelle classe d'objets , la classe *MRIdaggr*, qui va permettre d'agréger l'ensemble de l'information relative au patient :

- l'information clinique
exemple : âge, sexe, heure d'admission, ...
- les données d'imagerie
exemple : paramètres de diffusion, de perfusion, ...
- les données issues du pipeline de pré-traitement
exemple : type morphologique (CSF/WM/GM), hémisphère (gauche ou droit), ...
- les données obtenues lors de l'analyse
exemple : risque de nécrose prédit, position de la lésion, ...

1. données décrivant les données d'imagerie, comme les paramètres d'acquisition ou la résolution spatiale.

Librairie	Fonctionnalité(s) proposée(s)	Séquence(s) utilisée(s)	Format(s) reconnu(s)	Format(s) proposé(s)
oro.dicom	Lecture, écriture	non spécifique	fichier DICOM	
oro.nifti	Lecture, écriture	non spécifique	fichier NIFTI et ANALYSE	classe <code>nifti</code> ou <code>anlz</code>
tractor.base	Lecture, manipulation et visualisation	non spécifique	fichier DICOM, NIFTI et ANALYSE	classe <code>MriImage</code>
mrirc	Classification des voxels en CSF, WM ou GM	séquence T1	array	
dpmixsim	Segmentation d'images	non spécifique	vecteur	
brainR	Visualisation 3D ou 4D	non spécifique	fichier DICOM, NIFTI et ANALYSE	
EBImage	Traitement et segmentation d'images 2D		fichier jpeg, png et tiff	classe <code>Image</code>
mmand	Transformations morphologiques	non spécifique	array	
RNiftyReg	Recalage d'images	non spécifique	classe <code>nifti</code>	
fslr	Traitement et visualisation (interface avec le logiciel FSL)	non spécifique	fichier NIFTI et classe <code>nifti</code>	
dti	Analyse du tenseur en diffusion	séquence en diffusion	classe <code>dwi</code>	classe <code>dwi</code>
gdimap	Identification des fibres nerveuses	séquence en diffusion	fichier NIFTI	
dcemriS4	Calcul de l'ADC	séquence en diffusion	array	

TABLE III.2 – Synthèse des librairies R existantes pour l'analyse de données d'imagerie pour l'AVC. Sont exclues les librairies relatives à l'IRM fonctionnelle, à l'IRM à contraste amélioré dynamique, la tomographie par émission de positons et l'électroencéphalographie.



Cette classe dispose de fonctions spécifiquement écrites pour interagir avec elle, appelées méthodes. Elles vont faciliter l'extraction, le pré-traitement et la mise à jour des données présentes dans l'objet afin que la gestion soit la plus intuitive possible pour l'utilisateur. Pour cela, les commandes à écrire en code  sont proches du langage naturel.

Ainsi pour sélectionner le contraste du paramètre T1 pour les voxels de l'hémisphère gauche après soustraction de la valeur moyenne de l'hémisphère contralatéral, il suffit d'écrire :

```
1 R> selectContrast(objet,param="T1",hemisphere="left",norm_mu="contralateral")
```

Étant donné le nombre important de méthodes proposées (52 dans la version 1.1.2), le nom des méthodes a été standardisé selon leur rôle :

- méthodes facilitant la construction d'objets **MRIdaggr** (préfixe **const**)
exemple : **constMRIdaggr** permet de créer un objet **MRIdaggr** à partir d'une liste d'images IRM et d'un identifiant patient.
- méthodes de traitement d'images (préfixe **calc**)
exemple : **calcHemisphere** permet de définir, à partir d'un paramètre IRM, le plan sagittal séparant les hémisphères cérébraux.
- méthodes permettant d'accéder aux données contenues dans un objet **MRIdaggr**. Les données sont rangées dans des éléments typés appelés attributs.
exemple : l'attribut **contrast** contient les données d'imagerie. Elles peuvent être extraites avec la méthode **selectContrast**. Le type de cet attribut est un **data.frame** dont le nombre de lignes correspond au nombre de voxels de l'image.
- méthodes permettant d'ajouter ou de supprimer des données dans l'objet **MRIdaggr**. Afin de garantir la cohérence de l'objet, des tests de validité sont effectués lors de l'ajout de données. De plus, tous les attributs ne sont pas modifiables.
exemple : la méthode **allocContrast** permet d'ajouter un ou plusieurs paramètres d'imagerie. Cependant ceux-ci doivent fournir exactement une valeur par voxel pour que l'ajout soit effectif. Il n'est en revanche pas possible de modifier l'identifiant patient qui est fixé lors de la construction de l'objet.

La librairie est essentiellement écrite en langage  même si certaines fonctions de traitement d'images sont écrites en C++ pour gagner en temps de calcul. L'objectif n'étant pas de réécrire l'intégralité des outils de traitement d'images mais plutôt de combiner les ressources existantes, certaines fonctions servent essentiellement à rendre transparent pour l'utilisateur l'utilisation d'une fonction provenant d'une autre librairie. Ainsi **MRIdaggr** s'appuie sur différentes librairies (en plus des librairies de base du logiciel ) :

- *oro.dicom* et *oro.nifti* pour l'import et l'export de données.
- *spam* et *Matrix* pour l'utilisation de matrices de voisinage creuses.
- *mritc* pour la segmentation morphologique.
- *RANN* pour identifier le plus proche voisin d'un groupe spatial.

- *ROCR* pour le calcul de la courbe PR.
- *rgl* pour l'affichage graphique 3D de la lésion.
- *fields* pour l'affichage graphique 2D.
- *pracma* pour l'intégration numérique.

Une description plus exhaustive de l'objet `MRIdaggr` et des méthodes associées peut être trouvée dans l'article présenté en section 2.4. La librairie contient également une fonction pour calculer l'AUPRC avec son intervalle de confiance. L'estimation de l'intervalle de confiance utilise la méthode delta avec transformation logistique.

2.3 Pipeline de pré-traitements proposé

Dans cette section nous détaillons les méthodes et le paramétrage qui ont permis de préparer les données pour l'analyse.

La première étape du pipeline a été de normaliser le nom et l'arborescence des fichiers. À la suite de cette étape, les fichiers étaient triés par patient, puis par temps, puis par paramètre IRM. Pour assurer un import automatisé des données, les différents patients, temps et paramètres IRM étaient identifiés et une table répertoriant les données disponibles était créée.

La seconde étape a été l'import des fichiers images à l'aide de la fonction `readMRI` avec l'option `format="nifti"`, les fichiers étant en format NIFTI. Pour chaque patient, une liste était créée afin de stocker les images relatives à chacun des paramètres IRM. Puis un objet `MRIdaggr` était construit en utilisant la fonction `constMRIdaggr`. L'identifiant patient, le nom des paramètres et une liste contenant les images IRM étaient indiqués à la fonction. La résolution spatiale était automatiquement lue sur les métadonnées associées aux images par la fonction `constMRIdaggr`.

Une première série de traitements était ensuite appliquée sur l'objet :

- **uniformisation des résolutions spatiales** : quelques patients avaient une résolution de 256*256 pixels dans le plan. Leurs images étaient alors compressées à l'aide de la méthode `constCompressMRIdaggr` afin d'avoir une résolution commune de 128*128 pixels dans le plan pour l'ensemble de la cohorte. La résolution verticale était aussi variable (entre 13 et 20 coupes) mais elle n'a pas été modifiée.
- **incorporation des données cliniques** : à l'aide de la méthode `affectClinic`.
- **identification des groupes spatiaux de lésion** : afin de voir si la lésion était fragmentée ou non. La méthode `calcGroupsMask` était utilisée en fixant la distance maximale de voisinage à 6 mm (argument `W.range`) et en prenant en compte l'anisotropie de la résolution spatiale (argument `W.spatial_res`).
- **identification du cerveau** : à l'aide de la fonction `calcBrainMask`. Le paramètre de contraste utilisé était le T2 en écho de gradient. Un algorithme de k-moyennes avec entre

2 et 4 groupes était utilisé pour identifier le cerveau. Le résultat était ensuite retraité par `calcSmoothMask` pour améliorer la cohérence spatiale de la segmentation.

- **distribution de la lésion par coupe** : le nombre de voxels nécrosés sur chaque coupe était compté aux différents temps à l'aide de la méthode `calcTableLesion`. Cette méthode évalue le recouvrement de la lésion entre deux instants différents pour détecter d'éventuels problèmes de recalage (ou de détection de lésion). Elle permet aussi de s'assurer que le cerveau défini à l'étape précédente contient bien l'ensemble de la lésion.
- **restriction du domaine d'intérêt au cerveau** : en utilisant la méthode `constReduceMRIaggr` pour définir un objet `MRIaggr` réduit.

Enfin une seconde série de traitements était appliquée sur l'objet réduit :

- **identification des différents types morphologiques** : à l'aide de la méthode `calcTissueType`. Le paramètre T1 était utilisé ou, s'il n'était pas disponible, le T2 en écho de gradient.
- **identification des hémisphères** : en utilisant la méthode `calcHemisphere` avec le T1 ou à défaut le T2 en écho de gradient. L'initialisation était effectuée avec une recherche sur grille en utilisant 11 angles (argument `angle_test`) et 11 centres (argument `i_test`). Puis la position du plan sagittal médian retenue était optimisée avec un algorithme du simplex (argument `NelderMead`).
- **normalisation vis-à-vis du voxel controlatéral** : en utilisant le T1 ou à défaut le T2 pour associer les voxels de chaque hémisphère. La méthode `calcContralateral` était donc utilisée avec l'argument `type="1NN_penalised"`. La fenêtre de recherche était limitée à 3 voxels dans le plan avec un paramètre de pénalisation de 0,5 (argument `lambda`).
- **calcul de la matrice de voisinage** : la méthode `calcW` était utilisée avec une distance maximale de 10 mm (argument `range`), en prenant en compte l'anisotropie de la résolution spatiale (argument `spatial_res`). Une matrice de voisinage étant symétrique, seule la partie triangulaire supérieure était calculée et stockée dans l'objet (argument `upper`).
- **filtrage des paramètres IRM** : en utilisant la méthode `calcFilter` avec un filtre Gaussien (argument `filter="3D_G3"`) et un filtre bilatéral (argument `filter="w3D_G3"`).
- **normalisation intra-patient des paramètres IRM** : relativement à l'hémisphère controlatéral en utilisant la méthode `calcNormalization` avec l'argument `rm.CSF=FALSE` pour exclure le liquide cérébro-spinal.
- **définition de la distance à la lésion** : avec la fonction `calcDistMask` en prenant en compte l'anisotropie de la résolution spatiale (argument `spatial_res`).

Les résultats obtenus étaient satisfaisants excepté :

- pour les patients dont la boîte crânienne était proche du cerveau. Dans ce cas, la méthode d'identification du cerveau avait tendance à retenir des voxels de la boîte crânienne (voir l'image du milieu de la figure III.17). Les données étaient fournies avec une proposition de masque cérébral obtenu avec une méthode imposant un a priori de forme. Celle-ci était robuste vis-à-vis de la boîte crânienne mais donnait de mauvaises performances dans les coupes basses (voir l'image de droite de la figure III.17). Notre méthode tendait à fournir un masque cérébral parfois plus large (maximum +33,4%, minimum -9,5%, médiane -0,6%) mais qui recouvrait mieux la lésion (figure III.18).
- en cas de mauvais contraste en T1 (n=3) ou T2 (n=6), l'identification des types de tissus était insatisfaisante (figure III.19). Pour deux patients, de meilleurs résultats auraient pu être obtenus en utilisant une méthode de correction d'hétérogénéité du champ (ZHANG et al., 2001).

L'application du pipeline prenait en médiane 10 minutes par patient, principalement à cause de l'étape de segmentation du cerveau (3 minutes), d'identification des types de tissu (3 min) et de repérage des hémisphères (3 minutes).

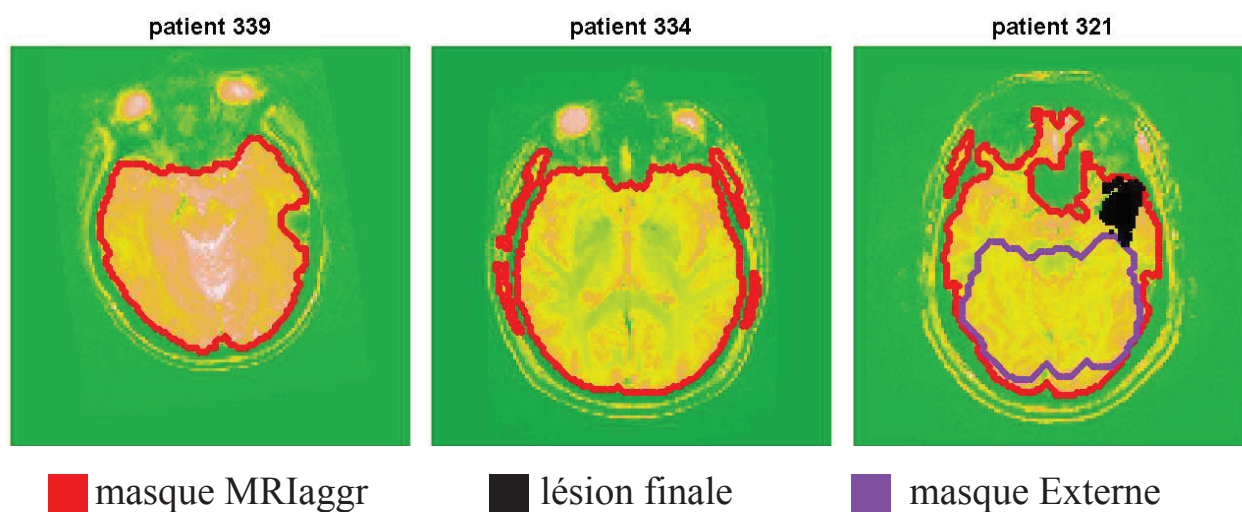


FIGURE III.17 – Exemple de masques cérébraux obtenus avec la librairie MRIaggr pour 3 patients (en rouge). À gauche, un masque cérébral satisfaisant. Au milieu, un masque cérébral incluant à tort des voxels de la boîte crânienne. À droite, le masque cérébral MRIaggr recouvre mieux la lésion (en noir) que la méthode imposant un a priori de forme (en violet).

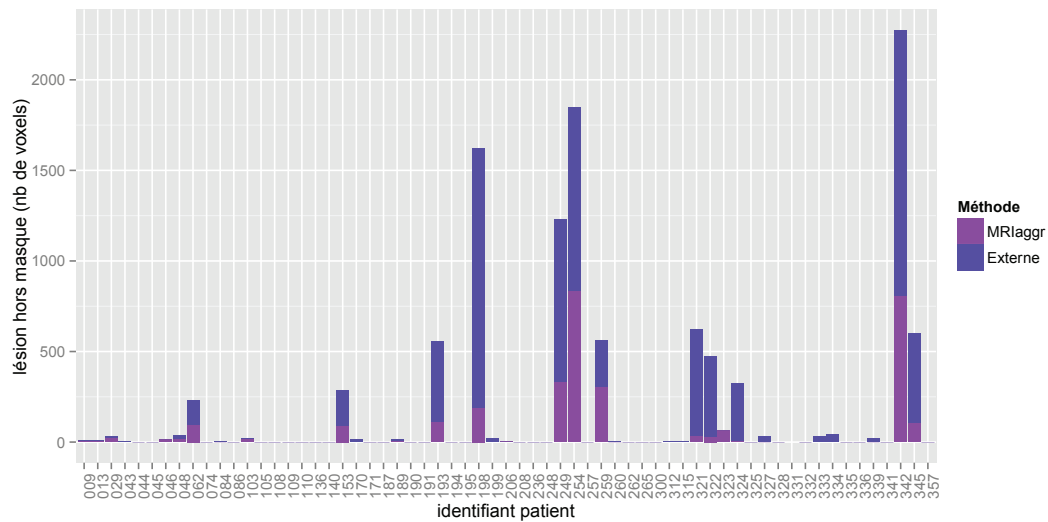


FIGURE III.18 – Nombre de voxels de la lésion finale hors du masque cérébral, défini soit en utilisant la librairie *MRIaggr* (en violet) soit avec une méthode imposant un a priori de forme (en bleu). L'idéal est la valeur 0 pour chaque patient puisque la lésion devrait être toute entière dans le masque cérébral.

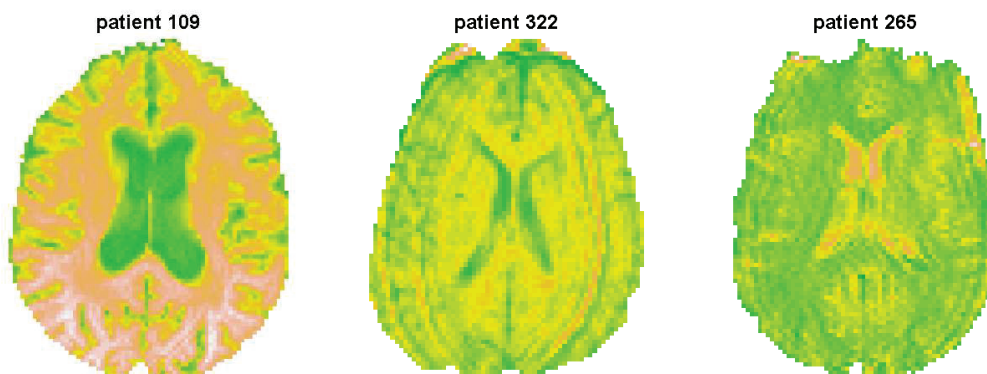


FIGURE III.19 – Exemple de contraste insuffisant pour l'identification du type de tissu. À gauche : T1 avec un gradient de contraste vertical dû à un problème d'homogénéité du champ magnétique lors de l'acquisition. Au milieu (respectivement à droite), contraste T1 (respectivement T2 en écho de gradient) insuffisant pour différencier la substance blanche de la substance grise, probablement lié à un mouvement du patient.

2.4 Article soumis au Journal of Statistical Software

Afin d'expliquer le fonctionnement de la librairie et d'améliorer sa visibilité, un article a été soumis au Journal of Statistical Software. De plus, la librairie contient trois fichiers de démonstration, une documentation associée à chaque fonction ainsi que des exemples d'utilisation.



Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II.

<http://www.jstatsoft.org/>

MRIaggr: An R Package for Multiparametric Medical Images

Brice Ozenne

Université Lyon 1
LBBE CNRS UMR 5558
Hospices Civils de Lyon

Delphine Maucourt-Boulch

Université Lyon 1
LBBE CNRS UMR 5558
Hospices Civils de Lyon

Abstract

The **MRIaggr** package for R software provides a convenient framework for image storage, processing, manipulation, and display. A new class, denoted **MRIaggr**, is able to aggregate data from several 3D imaging sequences into a single R object. Metadata can also be stored in this object.

The main functionalities of the package are organized in five categories associated with specific prefixes: construction (**const**), allocation (**alloc**), extraction (**select**), processing (**calc**), and displaying (**plot**). In particular, the package includes tools to import and export data from the **MRIaggr** object, which allows its use with standard statistical packages as well as other imaging packages. The processing functionalities include segmentation and morphological recognition tools designed for brain images.

The paper describes these functionalities according to the standard pipeline in image analysis, from importing data to displaying of the processing results. Illustrations are given at each step with MRI data on stroke.

Keywords: medical images, data storage, image processing, image display, R software.

1. Introduction

Medical imaging has enjoyed an extensive expansion during the last decades with the development of numerous sequences for both clinical diagnostics and medical research. In Magnetic Resonance Imaging (MRI), a large number of contrast parameters can be measured in each patient (PD, T1, T2, T2*, T2 FLAIR, DTI, DWI, ADC, CBF, CBV, MTT, TMAX, TTP, BOLD, etc.) making the management of multiparametric images from large cohorts a new challenge (Glasser *et al.* (2013)). Indeed, as soon as several patients are considered, the large amount of files complicates data management. Moreover, efficient functions are necessary

because imaging files contain large amounts of data –typically 512 x 512 voxels (volumetric picture element) per 2D slice; thus, about 10^7 voxels for a whole brain volume.

Despite the wide use of multiparametric imaging analysis (e.g., Wu *et al.* (2010), Reich *et al.* (2007), Kurhanewicza *et al.* (2008)), tools in R software are now available only for dealing with uniparametric imaging data (e.g., **tractor.base** package - Clayden *et al.* (2011)) possibly repeatedly measured in time (e.g., **fmri** package - Tabelow and Polzehl (2011)) but not with multiparametric imaging data (see Tabelow and Whitcher (2011) or <http://CRAN.R-project.org/view=MedicalImaging> for a comprehensive review on existing R packages). Besides, even when only one contrast parameter is of interest, the processing steps necessary to realign, normalize, or de-noise the images require often additional contrast parameters. A convenient framework for the management of multiparametric imaging data is thus clearly needed.

The proposed package, **MRIdaggr**, provides tools to import, extract, process, and display multiparametric medical images. It implements a new class of R objects, the **MRIdaggr** objects, that aggregate data from several 3D contrast parameters into a single object. Extraction, processing, and displaying functions are specifically designed to interact with this type of objects to improve data management. The article describes the main functionalities of the package according to the standard pipeline in image processing. The first section gives an overview of the package functionalities and presents the illustrative dataset used throughout this manuscript. The next section shows the way of importing and converting data into the **MRIdaggr** class. The remaining sections introduce the processing and displaying tools.

2. Package overview

2.1. Description of the main functionalities

This package is built around a new class of R objects called **MRIdaggr** that can store all data relative to a given patient (Figure 1). The package proposes functions specifically designed to interact with this type of objects (called methods). These functions may be divided into five categories:

- *Construction methods* that facilitate the creation of new **MRIdaggr** objects (prefix **const**).
- *Processing methods* that prepare the imaging data for analysis (prefix **calc**).
- *Extraction methods* that give access to the data present in a **MRIdaggr** object (prefix **select**).
- *Allocation methods* that update, add, or remove data in a **MRIdaggr** object (prefix **alloc** or **suppr**).
- *Displaying methods* that show the content of a **MRIdaggr** object into a graphical device (prefix or suffix **plot**).

NIFTI, Analyze, and DICOM imaging data can be imported into R using the **readMRI** function and converted into a **MRIdaggr** object with the **constMRIdaggr** function. Once created, several

processing methods can be applied to the `MRIaggr` object. For instance, the `calcBrainMask` method can be used to differentiate brain voxels from background voxels and the `calcFilter` method used to remove noise from the images. Some processing methods like `calcTissueType` or `calcDistMask` mainly interface `MRIaggr` objects to functions developed in other packages like `mrItc` (Feng and Tierney (2011)) or `RANN` (Arya *et al.* (2015)). If needed, the processing methods can directly store the results in the `MRIaggr` object.

Extraction methods give access to the data stored in a `MRIaggr` object. They include tools to

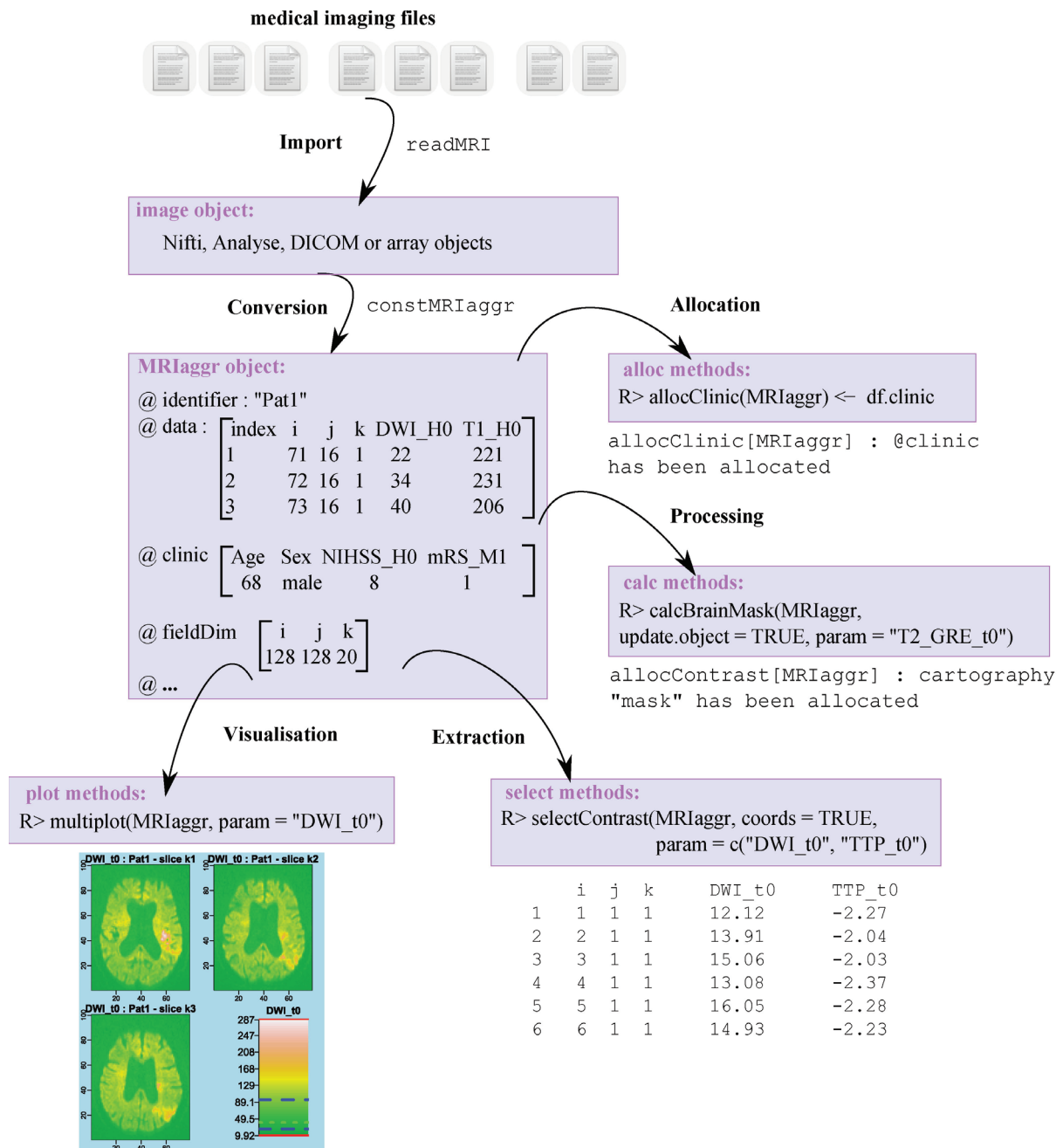


Figure 1: Overview of the functionalities of the `MRIaggr` package.

extract specific subsets of data (e.g., voxels from the third slice in the lesioned hemisphere), or to apply normalization operations (e.g., subtraction of the mean parameter value in the intact tissues). They can also give some information about the data available in the object. The modification of a **MRIaggr** object must be done using the allocation methods to ensure the consistency of the object. These methods can be used to remove unnecessary data from the object or to update the object with results from functions external to the package.

Finally, the package provides display tools. The main method is **multiplot**. It displays the values taken by a contrast parameter slice by slice. A set of points can be superimposed on the parameter map. When several patients are considered, the display of several contrast parameters generates quickly a large amount of files. Function **constLatex** has been developed to facilitate the display of these files by automatic generation of a **L^AT_EX** document structured by patient (section) and contrast parameter (subsection).

2.2. The example data

The functionalities of the package will be illustrated using the imaging data stored in the package. These data correspond to an extract from the MRI scans of a stroke patient. The initial 3D-images included 20 slices; only slices 13 to 15 are available in the package. The MRI images were acquired at three times: at hospital admission (t_0), after treatment (t_1), and one month later (t_2). At each time, several MRI parameters were measured:

t_0 : DWI, MTT, TTP, T1, T2 GRE, MASK DWI

t_1 : MTT, TTP, MASK DWI

t_2 : T2 FLAIR, MASK T2 FLAIR

DWI (Diffusion Weighted Imaging) and T2 FLAIR (Fluid Attenuated Inversion Recovery) are the gold standard MRI sequences for lesion identification. On these sequences, the lesion appears as a hyperintense area and was manually delineated by a physician. The resulting lesion indicator was stored under the name MASK DWI at t_0 and t_1 and MASK T2 FLAIR at t_2 . T1 and T2 GRE are morphological sequences, T1 is well suited for differentiating the white matter (WM) from the grey matter (GM) and T2 GRE gives a good contrast between brain voxels and background or skull voxels. Finally, the MTT (Mean Transit Time) and the TTP (Time To Peak) measure the quality of brain perfusion. Stroke is usually due to the occlusion of an artery, which delays the arrival of the blood to the tissue. MTT and TTP are different measures of this delay whose high values are synonymous of increased risk of necrosis (for a detailed introduction to MRI parameters within the context of acute stroke see [Muir et al. \(2006\)](#)).

To avoid confusing the names of MRI parameters measured at various times, these parameters will be denoted using the concatenation of their name and the observation time (e.g., T2 FLAIR t_2). In the R code, the spaces in the parameter names will be replaced by underscores (e.g., T2_FLAIR_t2).

3. Data importation, conversion, and storage

3.1. From the imaging files to the MRIaggr class

The NIFTI directory of the **MRIaggr** package contains the 11 NIFTI files that correspond to the 11 MRI parameters presented in Section 2.2. The path to this directory can be set using:

```
R> path <- system.file("nifti", package = "MRIaggr")
```

The `readMRI` function allows importing imaging files by giving the path and the format of the files. For NIFTI files, the function calls the `readNifTI` function of the **oro.nifti** package (Whitcher *et al.* (2011)) that imports the file in R under the form of a `nifti` object:

```
R> DWI_t0 <- readMRI(file = file.path(path, "DWI_t0.nii"), format = "nifti")
R> class(DWI_t0)
```

```
[1] "nifti"
attr(,"package")
[1] "oro.nifti"
```

```
R> DWI_t0
```

```
NifTI-1 format
```

```
  Type           : nifti
  Data Type      : 16 (FLOAT32)
  Bits per Pixel : 32
  Slice Code     : 0 (Unknown)
  Intent Code    : 0 (None)
  Qform Code     : 0 (Unknown)
  Sform Code     : 0 (Unknown)
  Dimension      : 78 x 100 x 3 x 1
  Pixel Dimension : 0 x 0 x 0 x 0
  Voxel Units    : Unknown
  Time Units     : Unknown
```

The `constMRIaggr` function facilitates the conversion from `nifti` objects to **MRIaggr** objects. This function has, as main argument, the imaging data to store. This argument should be passed to the function using a list of `nifti` objects, each `nifti` object corresponding to a contrast parameter. To create this list, a `for` loop over the contrast parameters can be used:

```
R> MRIparameters <- c("DWI_t0", "TTP_t0", "MTT_t0", "T1_t0", "T2_GRE_t0",
+   "MASK_DWI_t0", "TTP_t1", "MTT_t1", "T2_FLAIR_t2", "MASK_T2_FLAIR_t2")
R> list.nifti <- list()
R> for(iter_param in 1:length(MRIparameters)){
+   iter_file <- file.path(path, MRIparameters[iter_param])
+   list.nifti[[iter_param]] <- readMRI(file = iter_file, format = "nifti")
+ }
```

Then the `constMRIaggr` can be called by specifying the identifier of the patient, a vector containing the names of each contrast parameter and, if not present in the `nifti` object, the voxel dimensions:

```
R> voxelDim <- data.frame(i = 1.875, j = 1.875, k = 6, unit = "mm")
```

```
R> MRIdaggr.Pat1 <- constMRIdaggr(list.nifti, identifier = "Pat1",
+   param = MRIparameters, voxelDim = voxelDim)
```

```
Merging : (1) DWI_t0 (2) TTP_t0 (3) MTT_t0 (4) T1_t0 (5) T2_GRE_t0
(6) MASK_DWI_t0 (7) TTP_t1 (8) MTT_t1 (9) T2_FLAIR_t2 (10) MASK_T2_FLAIR_t2
```

This function extracts first the contrast parameter values from each `nifti` object then converts these values from the wide format (where the coordinates are given by their positions in the array) to the long format (where values are stored into a `data.frame` with one line per voxel and three columns that contain voxel coordinates). After checking the consistency of the coordinates between the `data.frames`, the function merges all the data into a single `MRIdaggr` object.

Importation of Analyze files and DICOM files can be done in a similar way, e.g:

```
R> path.Pat2 <- system.file("anlz", package = "oro.nifti")
R> analyze.avg <- readMRI(file.path(path.Pat2, "avg152T1"),
+   format = "analyze")
R> MRIdaggr.Pat2 <- constMRIdaggr(analyze.avg, param = "T1_t0",
+   identifier = "Pat2")
```

```
Merging : (1) T1_t0
```

However, for imaging data acquired repeatedly in time (i.e., 4-dimensional data), one parameter must be passed at a time:

```
R> path.Pat3 <- system.file("nifti", package = "oro.nifti")
R> nifti.ffd <- readMRI(file.path(path.Pat3, "filtered_func_data"),
+   format = "nifti")
R> MRIdaggr.Pat3 <- constMRIdaggr(list(nifti.ffd[,,,1,drop = FALSE],
+   nifti.ffd[,,,2,drop = FALSE], nifti.ffd[,,,3,drop = FALSE]),
+   param = c("T2_GRE_t0", "T2_GRE_t1", "T2_GRE_t2"), identifier = "Pat3")
```

```
Merging : (1) T2_GRE_t0 (2) T2_GRE_t1 (3) T2_GRE_t2
```

3.2. Structure of the `MRIdaggr` class

The `MRIdaggr` package uses the S4 classes and methods defined in R software (for a presentation of the S4 class, see chapters 9 and 10 of [Chambers \(2008\)](#)). Readers unfamiliar with the concept of class can see `MRIdaggr` objects as a list of several elements (called attributes) where each element has a pre-defined type (e.g., matrix). Each `MRIdaggr` object is composed of the following attributes:

- *identifier*: a character identifying the patient.
- *contrast*: a `data.frame` containing the contrast parameter values. Parameter values are stored in columns and voxel values in lines.
- *clinic*: the clinical data of the patient.
- *fieldDim*: the dimensions of the spatial field containing the voxels.
- *voxelDim*: the voxel dimensions.

- *default_value*: a reference value for each contrast parameter.
- *history*: a list of the processing steps that have been applied to the object.
- *normalization*: the normalization values of each contrast parameter.
- *hemispheres*: the presence or absence of lesion(s) in each hemisphere.
- *midplane*: the position of the mid-sagittal plane.
- *table_lesion*, *table_hypo*, *table_reperf*: volumic information about the lesion, the hypoperfusion area, and the reperfusion area.
- *ls_descStats*: a list containing additional data. There is no pre-defined type for the elements of this list; any R result can be stored here.

The `constMRIaggr` function will fill the main attributes of the `MRIaggr` object: *identifier*, *contrast*, *fieldDim*, *voxelDim*, and *default_value*. The other attributes will be computed during the processing stage. Coming back to the example, the content of the object can be displayed using the method `summary`:

```
R> summary(MRIaggr.Pat1)
```

```
Object of class 'MRIaggr' with identifier : Pat1
# image dimensions (i,j,k) : 78x100x3 voxels
# voxel dimensions (i,j,k unit) : 1.875x1.875x6 mm
# history : 0 calculations performed on the object
# number of observations : 23400
# number of contrast parameters : 10
# tables : lesion empty, hypoperfusion empty, reperfusion empty
# normalization values : empty
# hemisphere left : undefined | hemisphere right : undefined
# midplane coordinates : empty
# ls_descStats : empty
# clinic : empty
```

3.3. Data management with a `MRIaggr` object

The user accesses the attributes of the object via extracting methods. By default these methods return the entire attribute but it is also possible to request only a subset of this attribute by using the arguments of the method. For instance `selectContrast` can extract all the contrast parameter values:

```
R> res <- selectContrast(MRIaggr.Pat1)
```

```
R> str(res)
```

```
'data.frame':      23400 obs. of  10 variables:
 $ DWI_t0      : num  12.1 13.9 15.1 13.1 16 ...
 $ TTP_t0      : num  -2.27 -2.04 -2.03 -2.37 -2.28 ...
 $ MTT_t0      : num  -4.36 -4.34 -4.41 -4.48 -4.53 ...
 $ T1_t0       : num   5.07 5.86 3.86 6.09 7.98 ...
 $ T2_GRE_t0   : num   6.08 4.92 7.96 6.98 6.93 ...
```


8

***MRIdaggr**: Management of multiparametric images*

```

$ MASK_DWI_t0      : num  0 0 0 0 0 0 0 0 0 0 ...
$ TTP_t1           : num  -3.42 -3.28 -3.26 -3.31 -3.51 ...
$ MTT_t1           : num  -10.4 -10.3 -10.5 -10.6 -10.4 ...
$ T2_FLAIR_t2      : num   6.81 9.95 8.9 8.29 6.89 ...
$ MASK_T2_FLAIR_t2: num   0 0 0 0 0 0 0 0 0 0 ...

```

`selectContrast` can also extract the DWI t0 values of the second slice by specifying the `num` and the `param` argument:

```

R> res <- selectContrast(MRIdaggr.Pat1, num = 2, param = "DWI_t0")
R> str(res)

```

```

'data.frame':      7800 obs. of  1 variable:
 $ DWI_t0: num  13 12.1 15.1 12.9 12 ...

```

The extraction of the spatial coordinates is performed by an additional extractor named `selectCoords`:

```

R> res <- selectCoords(MRIdaggr.Pat1)
R> str(res)

'data.frame':      23400 obs. of  3 variables:
 $ i: num   1 2 3 4 5 6 7 8 9 10 ...
 $ j: num   1 1 1 1 1 1 1 1 1 1 ...
 $ k: num   1 1 1 1 1 1 1 1 1 1 ...

```

Elements can be added or removed from the `MRIdaggr` object using the allocation methods, each allocation method being relative to a single attribute. However, the new value must meet some validity conditions to ensure the consistency of the object. For instance, the `allocContrast` method can be used to allocate a new contrast parameter:

```

R> n.voxels <- selectN(MRIdaggr.Pat1)
R> allocContrast(MRIdaggr.Pat1, param = "newP") <- rnorm(n.voxels)

allocContrast[MRIdaggr] : Cartography "newP"
                        has been allocated

```

where the `selectN` method returns the total number of voxels. After checking the compatibility between the number of voxels in the object and the length of the new parameter, a new parameter "newP" is allocated to the object. If the new value does not pass the checks, the update of the object is canceled and an error message is returned:

```

R> allocContrast(MRIdaggr.Pat1, param = "newP") <- rnorm(5)

Error in `allocContrast<-`(`*tmp*`, param = "newP", value = c(1.10650817970865,
allocContrast[MRIdaggr] : mismatch between the dimension of 'value' and
'object@contrast'
nrow(object@contrast) : 23400
nrow(value) : 5

```

If an element of the object has to be changed, argument `overwrite` has to be set to `TRUE`. This is a security to avoid unintentional overwriting.

```

R> allocContrast(MRIdaggr.Pat1, param = "newP", overwrite = TRUE) <- rep(1,
+   n.voxels)

```



```
allocContrast[MRIaggr] : Cartography "newP"
                        has been updated
```

Contrast parameters can be removed using `supprContrast`:

```
R> supprContrast(MRIaggr.Pat1) <- "newP"

supprContrast[MRIaggr] : Cartography "newP"
                        has been removed
```

The user is allowed to change all attributes except `identif`, `fieldDim`, `voxelDim`, `default_val` and `history`. The first four are defined at the construction of the object and there is no reason to change them afterwards. The latter will automatically register the arguments passed to any processing method that changed the object.

3.4. Interaction with external functions

Most functions of standard statistical packages recognize only the long format. Because this format is the one in which the imaging data are stored and exported, these functions can be conveniently used in interaction with `MRIaggr` objects:

```
R> data.df <- selectContrast(MRIaggr.Pat1, param = c("DWI_t0", "MASK_DWI_t0"))
R> coords.df <- selectCoords(MRIaggr.Pat1)
R> head(cbind(coords.df, data.df))
```

	i	j	k	DWI_t0	MASK_DWI_t0
1	1	1	1	12.12	0
2	2	2	1	13.91	0
3	3	3	1	15.06	0
4	4	4	1	13.08	0
5	5	5	1	16.05	0
6	6	6	1	14.93	0

```
R> glm_DWI <- glm(MASK_DWI_t0 ~ DWI_t0, data = data.df)
```

However, packages specialized in image processing like **mmand** (Clayden (2014)) or **EBImage** (Pau *et al.* (2010)) recognize often only the wide format. In such a case, a conversion from the long format to the wide format can be made using the `df2array` function. This function requires the parameter values and the voxel coordinates to be set:

```
R> data.array <- df2array(contrast = data.df, coords = coords.df)
```

The output of `df2array` contains, in its first element named `contrast`, the list of the parameter values stored in the array format:

```
R> lapply(data.array[["contrast"]], dim)
```

```
$DWI_t0
  i  j  k
78 100 3
```

```
$MASK_DWI_t0
  i  j  k
78 100 3
```

Note that the resulting array is equivalent to the original array, except that, in the original

array, the data were stored in four dimensions:

```
R> quantile(data.array[["contrast"]][["DWI_t0"]]-DWI_t0[,,,1])
 0%  25%  50%  75% 100%
 0    0    0    0    0
```

An inverse symmetrical image can be performed using the `flip` function of the `EBImage` package:

```
R> require(EBImage)
R> data.array_flip <- flip(data.array[["contrast"]][["DWI_t0"]])
```

and the reverse conversion can be performed using the `array2df` function:

```
R> data.df_flip <- array2df(data.array_flip)
```

The result can then be stored into the `MRIdaggr` object:

```
R> allocContrast(MRIdaggr.Pat1, param = "DWI_t0_flip") <- data.df_flip
allocContrast[MRIdaggr] : Cartography "DWI_t0_flip"
                        has been allocated
```

4. Processing tools

The analysis of medical imaging data requires several preliminary processing steps that lead to workable data. Several methods are available for this purpose (see Table 1); their main characteristics are presented in the following subsections. The first argument of all the methods is the `MRIdaggr` object with which they will interact.

Background removal	
<code>calcBrainMask</code>	Brain segmentation using a thresholding or a k-means approach
<code>calcSmoothMask</code>	Spatial smoothing of the brain mask
<code>constReduceMRIdaggr</code>	Generation of a <code>MRIdaggr</code> object restricted to a subset of voxels
Morphological segmentation	
<code>calcHemisphere</code>	Identification of the mid-sagittal plane
<code>calcTissueType</code>	CSF, GM, and WM probabilistic segmentation
Contrast parameter processing	
<code>calcFilter</code>	Contrast parameter filtering
<code>calcRegionalIntensity</code>	Regional mean of the contrast parameter value
<code>calcThresholdMRIdaggr</code>	Contrast parameter thresholding
Contrast parameter normalization	
<code>calcContralateral</code>	Normalization relatively to contralateral voxels
<code>calcNormalization</code>	Normalization relatively to the contralateral CSF, GM, or WM reg

Table 1: Processing methods for `MRIdaggr` objects. CSF=cerebro-spinal fluid, GM=grey matter, WM=white matter.

4.1. Background removal

Medical images are usually acquired with a larger field of view than the region of interest. Thus, a background removal step is required to isolate the relevant region (e.g., brain tissue). The difficulty of the operation depends on the contrast of the image and on the geometry of the region of interest. Several approaches have been developed (for a comprehensive review see Bankman *et al.* (2009)); some have been implemented in specific programs (e.g., FreeSurfer software <http://surfer.nmr.mgh.harvard.edu/>) but they are not available in R.

The `calcBrainMask` method implements two classical approaches: thresholding and k-means partition. By analyzing the intensity histogram of a contrast parameter, each approach aims to find the mask of the region of interest giving value one inside the region and zero outside it. Argument `type` determines the approach to be used ("`threshold`" or "`k-means`") and the contrast parameter is set by argument `param`. The latter has to correspond to the name of a contrast parameter stored in the `MRIaggr` object.

Thresholding is the elementary way to segment an image. It attributes a voxel to the mask if the contrast value of this voxel exceeds a pre-defined threshold. One drawback of this approach is that the optimal threshold can be difficult to find and may vary from one image to another. To help the user choosing the threshold, the proposed method computes the number of voxels belonging to the mask for a variety of thresholds. The extrema of the derivative of this function should correspond to the removal of a new anatomical structure. Then, instead of passing the threshold value to the method, the user has to specify which optimum is intended to be used via argument `th.select_optima`:

```
R> res4_a1 <- calcBrainMask(MRIaggr.Pat1, param = "T2_GRE_t0",
+   type = "threshold", th.select_optima = 2)
```

Derivative has been smoothed with a Gaussian kernel of width 5 breaks

Threshold : derivative (selected)

10.16	:	1440.8	
176.48	:	19.375	(*)
252.08	:	711.31	
383.12	:	79.312	
428.48	:	249.28	

As stated by the output, the derivative is smoothed to remove non-significant optima. The (*) indicates the threshold that corresponds to the optimum selected by the user (here 176.48). The method returns also a graphical representation of the number of voxels that belongs to the brain mask and its derivative according to the threshold (see Figure 2).

The k-means approach partitions the histogram of the contrast parameter with different numbers of groups and selects the partition with the highest spatial coherence. Given a maximum number of groups G , the algorithm performs a k-means partition of the histogram for each number of group $k \in [2, G]$ and forms several binary partitions from this partition (e.g., when $k = 3$: group 1 vs. group 2 and 3, or group 1 and 2 vs. group 3). Then it assesses the spatial coherence of each partition by counting over the mask voxels the average number of neighbors that belongs to the mask:

$$V(mask) = \frac{1}{\sum_{i \in mask} \sum_{j \in N(i)} 1} \sum_{i \in mask} \sum_{j \in N(i)} \mathbb{1}_{j \in mask}$$

where $N(i)$ denotes the neighborhood of voxel i and $\mathbf{1}_x$ the indicator variable that takes value 1 if x is true, otherwise 0. The only argument required for the k-means approach is the maximum number of groups per partition (`kmeans.n_groups`, denoted G above). For a brain MRI T2 sequence, G should be at least 4 to allow for background, cerebro-spinal fluid (CSF), WM, and GM:

```
R> res4_a2 <- calcBrainMask(MRIaggr.Pat1, param = "T2_GRE_t0",
+   kmeans.n_groups = 2:4, update.object = TRUE)

1 brain groups : G2.=.2
2 brain groups : G3.=.2 G3.=.3 G3.>=.2
3 brain groups : G4.=.2 G4.=.3 G4.=.4 G4.>=.3 G4.<=.3
allocContrast[MRIaggr] : Cartography "mask"
                        has been allocated
```

The method indicates the tested partitions. For instance, `G4.>=.3` is the result of the k-means algorithm with 4 groups of whom the 2 groups with the lowest contrasts (groups 1 and 2) are associated with the background. Setting `update.object` to `TRUE` leads to an automatic update of the object. Indeed, the output indicates that a contrast parameter named "mask" has been allocated to the object. This parameter corresponds to the best partition and equals 0 for background voxels and 1 for brain voxels.

One major drawback of thresholding and k-means is that none imposes spatial regularity to the group shape. To correct this, the partition can be smoothed with method `calcSmoothMask`. This method fills the holes of the mask and erases isolated voxels:

```
R> res4_a3 <- calcSmoothMask(MRIaggr.Pat1, update.object = TRUE, overwrite = TRUE)

rm small2D : *****
  14 vx ; 14 groups
rm small3D : *****
```

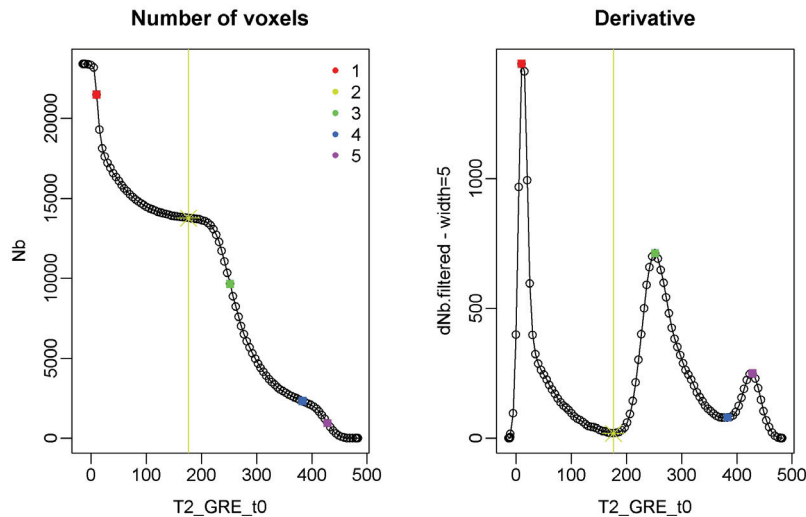


Figure 2: Number of voxels present in the brain mask according to the thresholding method (left graph) and the smoothed derivative (right graph). In both graphs, the extrema are displayed as colored rectangles. The selected extremum is the one crossed by the vertical line.

```

0 vx ; 0 groups
erosion : *****
0 vx ; 0 groups
add hole3D : *****
3 vx ; 3 groups
smoothing (add/rm) : (1) 0/1 (2) 0/1 (3) 0/0
allocContrast[MRIaggr] : Cartography "mask"
                        has been updated

```

By default, in the object, this method will smooth the contrast parameter named "mask" (another contrast parameter can be smoothed by passing its name to argument `mask`). Figure 3 displays one slice of each of the previous masks. The panels of this figure were obtained using the display method `multiplot` that will be presented in Section 5. In this example, k-means and thresholding approaches led to close results and the spatial processing of the mask led to a satisfactory result.

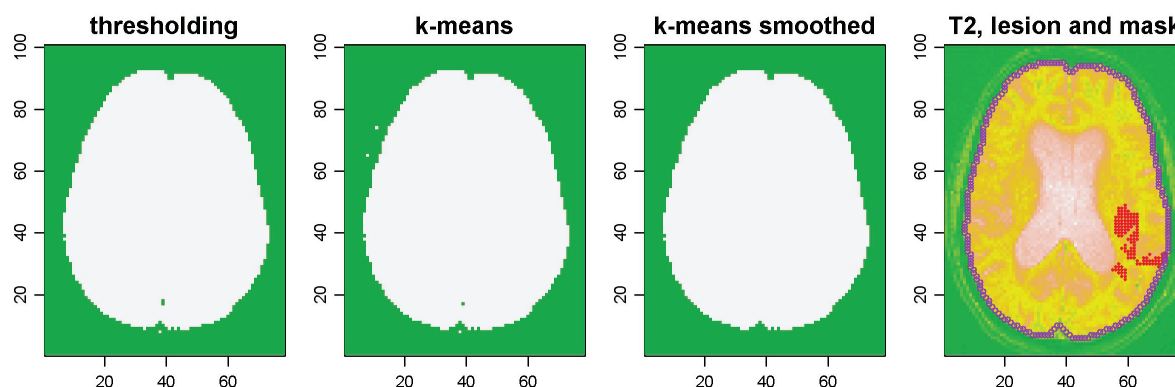


Figure 3: From left to right: Brain masks obtained after thresholding, after k-means, after k-means and smoothing (brain voxels appear in white and background voxels in green), and outline of the smoothed mask (in purple) and the lesion (in red) on the T2 GRE t0 contrast map.

Because background voxels are often of no interest for the analysis and increase the need for memory resources, it may be interesting to exclude them. The `constReduceMRIaggr` method enables the construction of a reduced `MRIaggr` object in which the background voxels are removed from attribute `contrast` according to a binary indicator (argument `mask`):

```

R> MRIaggr.Pat1_red <- constReduceMRIaggr(MRIaggr.Pat1, mask = "mask")
supprContrast[MRIaggr] : Cartography "mask"
                        has been removed
allocDescStats[MRIaggr] : Element "index_sauve"
                        has been allocated
constReduceMRIaggr[MRIaggr] : MRIaggr_red has been created
R> selectN(MRIaggr.Pat1_red)
[1] 13799
R> object.size(MRIaggr.Pat1) / object.size(MRIaggr.Pat1_red)

```

14

MRIaggr: Management of multiparametric images

1.59442395040064 bytes

The number of voxels included in the object decreased from 23400 to 13799 and the size of the object was approximately divided by 1.6.

4.2. Morphological segmentation

Identifying the morphological brain structures may have several purposes such as the stratification of the analysis on the morphological structures or the normalization of the contrast parameters relatively to values from a morphological structure (e.g., the mean contrast value of the contralateral WM). The **MRIaggr** package provides two structural recognition methods: **calcHemisphere** for the mid-sagittal plane identification and **calcTissueType** for CSF, GM, and WM segmentation.

The mid-sagittal plane identification relies on the maximisation of a symmetry criterion to find the optimal plane that separates the brain into two symmetrical regions. Assuming perfect registration between slices, the plane intersects each slice following the same line segment. Then the search is reduced to two degrees of freedom: one for the abscissa of the center of the line segment and one for the orientation of the line segment. Then, for each pair of parameters, the following symmetry criterion is computed:

$$\sum_{i=1}^n \frac{1}{1 + |X(i) - X_{contro}(i)|^p / \sigma^p}$$

where $X(i)$ is the contrast of voxel i
 $X_{contro}(i)$ is the contralateral contrast of voxel i
 σ is the standard deviation of the contrast parameter
 p is a pre-defined parameter (by default $p = 2$)

and the pair leading to the maximum value is retained. When the symmetric of a voxel i with respect to the mid-sagittal plane overlaps several voxels, the contrasts of the overlapping voxels are averaged with weights that correspond to the amount of overlap to provide the contralateral contrast. The call of this method must specify the contrast parameter to use. Optionally, argument **mask** can also specify one or several logical contrast parameters to indicate the position of the lesion. This allows distinguishing the lesioned from the intact hemisphere.

```
R> calcHemisphere(MRIaggr.Pat1_red, param = "T2_GRE_t0",
+ mask = "MASK_T2_FLAIR_t2", as.logical = TRUE, update.object = TRUE)

Grid Search : 117 parametrisations
i      : -20 -15 -10 -5 0 5 10 15 20 (in voxels)
angle  : -30 -25 -20 -15 -10 -5 0 5 10 15 20 25 30 (in degrees)
***** cv
Nelder-Mead with optim
allocContrast[MRIaggr] : Cartographies "i_hemisphere" "j_hemisphere" "hemisphere
                        have been allocated
allocHemisphere[MRIaggr] : @midplane has been allocated
allocHemisphere[MRIaggr] : @hemispheres has been allocated
```

As shown by the output, the position of each voxel with respect to the mid-sagittal plane, the position of the mid-sagittal plane, and the presence of lesions in the hemispheres have

been stored in the object. Due to the large amount of computation, this method can be slow (approximately 20 seconds in this example). Restricting the size of the grid search (argument `i_test` and `angle_test`), decreasing the maximum number of iterations of the Nelder-Mead optimizer (argument `maxit`), or evaluating the symmetry criteria on a subset of slices (argument `num`) can reduce the computation time at the cost of a less accurate estimation. The estimated mid-sagittal plane is displayed on Figure 4 (left panel).

`calcTissueType` is an interface to the `mrirc.bayes` function of the `mrirc` package. This function performs a probabilistic segmentation of brain voxels into CSF, GM, and WM classes. It relies on a spatial finite mixture model that is a compromise between grouping voxels with close contrasts and favoring the spatial homogeneity of the groups (see [Feng et al. \(2012\)](#) for a comprehensive presentation of the method). The method uses two arguments: the choice of the contrast parameter and the value of the parameter that favors spatial homogeneity. T1 contrast is the recommended sequence to use and, by default, the spatial homogeneity parameter is set to a pre-defined value:

```
R> calcTissueType(MRIaggr.Pat1_red, param = "T1_t0", update.object = TRUE)

10% of iterations has finished
20% of iterations has finished
30% of iterations has finished
40% of iterations has finished
50% of iterations has finished
60% of iterations has finished
70% of iterations has finished
80% of iterations has finished
90% of iterations has finished
100% of iterations has finished
allocContrast[MRIaggr] : Cartographies "CSF" "GM" "WM"
                        have been allocated
```

Three new contrast parameters that represent the probabilities of membership of each voxel to the CSF, the GM, or the WM have been stored in the object. Figure 4 (right panel) displays the probabilities of membership using mixtures of red, green, and blue. For example, a voxel that belongs to the CSF with probability 0.61, to the WM with probability 0.24 and, to the GM with probability 0.01 is displayed using a mixture of 61% red, 24% green, and 1% blue.

4.3. Contrast parameter processing

Three methods are available to process the contrast parameters. The first one, `calcRegionalMRI`, computes the regional mean value of a contrast parameter using a Gaussian kernel to weight neighboring values:

$$X_{regional} = \sum_{j \in N(i)} w_{ij} X(j) \text{ where } w_{ij} = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{dist(i,j)^2}{\sigma^2}\right)$$

For this, a neighborhood matrix is needed. If this matrix is not passed to `calcRegionalMRI` through argument `W`, it will be automatically computed using the voxel coordinates. Because the computation of the entire neighborhood matrix would be expensive and unnecessary, only

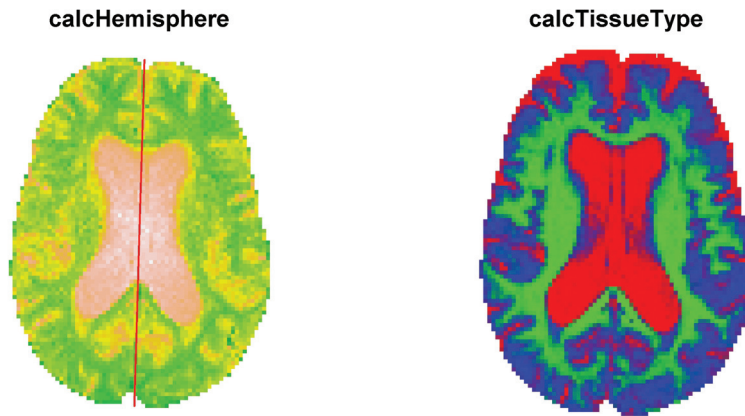


Figure 4: The mid-sagittal plane (left panel, red line) and the probabilities of membership of each voxel to the CSF (red), the WM (green), and the GM (blue) (right panel) as displayed by the `multiplot` function.

distances below the value of the `W.range` argument are recorded. Additionally, in case of non-isotropic voxel resolution, argument `W.spatial_res` enables specifying the voxel dimension in each direction. The bandwidth of the kernel has to be specified using argument `bandwidth`. When applying the method on several contrast parameters, the same bandwidth is used for all parameters:

```
R> calcRegionalContrast(MRIaggr.Pat1_red, param = c("T2_FLAIR_t2", "T1_t0"),
+   bandwidth = 1.875, W.spatial_res = c(1.875, 1.875, 6),
+   W.range = 6, update.object = TRUE)

computing W ... W ready
param : T2_FLAIR_t2 T1_t0
allocContrast[MRIaggr] : Cartographies "T2_FLAIR_t2_regional" "T1_t0_regional"
                        have been allocated
```

Filtering operations can be performed with method `calcFilter`. A filtering operation is defined by a range and an operator: for each voxel, the coordinate system is centered on this voxel and the operator is applied to all voxels within the range. The method accepts either user-defined filters or names of standard filters. User-defined filters must be arrays whose dimensions define the range and whose values define the weights attributed to various positions:

```
R> Gfilter <- array(dbinom(0:2, 2, 0.5), dim = 3) %o% array(dbinom(0:2, 2, 0.5),
+   dim = 3) %o% array(dbinom(0:2, 2, 0.5), dim = 3)
R> resfilter1 <- calcFilter(MRIaggr.Pat1_red,
+   param = c("DWI_t0", "TTP_t0"), filter = Gfilter)$res
DWI_t0 TTP_t0
```

The filtered value is then a weighting sum of the neighboring contrast values:

```
R> coords <- resfilter1[,c("i", "j", "k")]
R> DWI_t0.array <- df2array(selectContrast(MRIaggr.Pat1_red,
```

```

+ param = "DWI_t0"), coords = coords,
+ range.coords = c(78, 100, 3))$contrast[[1]]
R> sum(DWI_t0.array[51:53, 51:53, 1:3] * Gfilter)
[1] 96.052

R> indexVx <- which(apply(coords, 1, function(x){sum(x == c(52, 52, 2))}) == 3)
R> resfilter1[indexVx, "DWI_t0_filtered"]
[1] 96.052

```

The standard filters have to be called using a character string. The first two characters specify whether filtering should be performed in 2D or 3D. The third character must be an underscore. The fourth character specifies the filter:

- **G** for Gaussian filter that computes the voxel weights using a Gaussian kernel on the distance to the central voxel.
- **M** for median filter that computes the median of the values.
- **S** for Sobel filter that computes the gradient in direction i, j, or k.
- **I** for influence filter that sums all the values.

And the last character indicates the range of the filter (except for the Sobel filter where it indicates the direction). Then a 3D Gaussian smoothing with a 3-voxel range can be called using `filter = "3D_G3"`:

```

R> resfilter2 <- calcFilter(MRIaggr.Pat1_red, param = c("DWI_t0", "TTP_t0"),
+ filter = "3D_G3", update.object = TRUE)
DWI_t0 TTP_t0
allocContrast[MRIaggr] : Cartographies "DWI_t0_3D_G3" "TTP_t0_3D_G3"
                        have been allocated

```

In this example, both approaches specified the same filter:

```

R> quantile(Gfilter-resfilter2[["filter"]])
0% 25% 50% 75% 100%
0 0 0 0 0

```

so, the output was the same:

```

R> quantile(resfilter1[, "DWI_t0_filtered"]
+ - resfilter2[["res"]][, "DWI_t0_3D_G3"])
0% 25% 50% 75% 100%
0 0 0 0 0

```

Edge preserving filters can be obtained by setting argument `w_intensity` to `TRUE` (not available for median filters). For a given voxel i , this will lead to weight each neighboring voxel j by the following weight:

$$w_j = \frac{f_{\mathcal{N}}(X_j - X_i, \sigma_X)}{f_{\mathcal{N}}(0, \sigma_X)}$$

where σ_X denotes the standard deviation of the contrast parameter X and $f_{\mathcal{N}}$ the density of a normal law. Thus, neighboring voxels with the same intensity as the central voxel will

have a dominant weight: neighboring voxels from one side of an edge will not significantly influence the filtering of a voxel from the other side (Tomasi and Manduchi (1998)).

Figure 5 displays a comparison between the initial contrast parameter and the contrast obtained after application of different filters:

```
R> calcFilter(MRIaggr.Pat1_red, param = "DWI_t0",
+   w_contrast = TRUE, name_newparam = "DWI_t0_w3D_G3",
+   filter = "3D_G3", update.object = TRUE, trace = FALSE)
R> calcFilter(MRIaggr.Pat1_red, param = "DWI_t0",
+   filter = "3D_M3", update.object = TRUE, trace = FALSE)
R> calcFilter(MRIaggr.Pat1_red, param = "DWI_t0", na.rm = TRUE,
+   filter = "2D_Sx", update.object = TRUE, trace = FALSE)
R> calcFilter(MRIaggr.Pat1_red, param = "DWI_t0", na.rm = TRUE,
+   w_contrast = TRUE, name_newparam = "DWI_t0_w2D_Sx",
+   filter = "2D_Sx", update.object = TRUE, trace = FALSE)
```

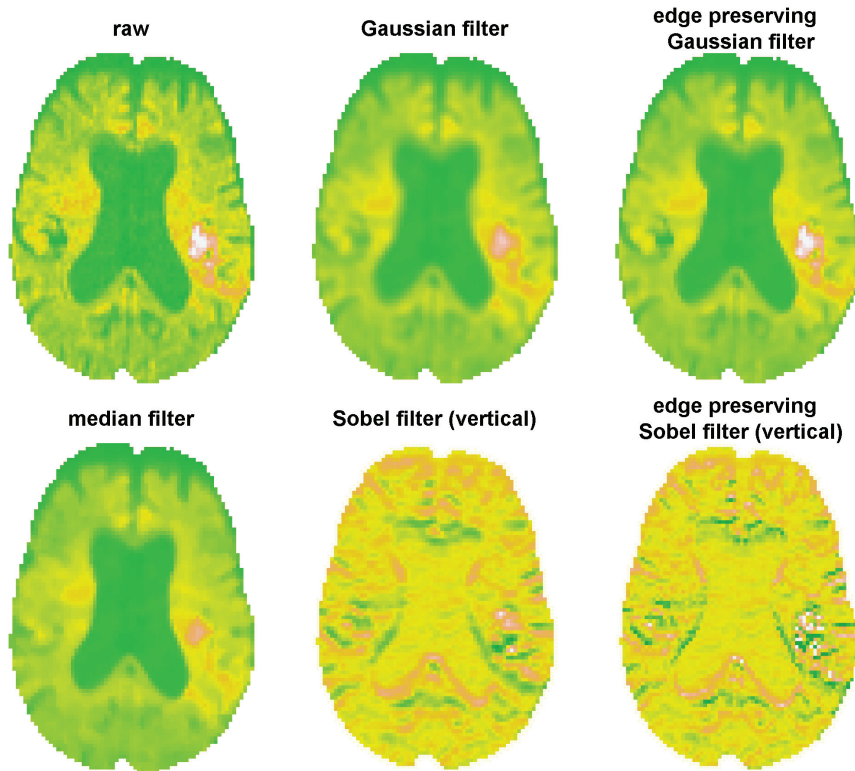


Figure 5: Initial contrast parameter (top left panel) and contrast parameters after application of different filters (other panels).

The `calcThresholdMRIaggr` is dedicated to the computation of thresholded maps. Such maps are used to identify at-risk tissue at a given moment of acute stroke (e.g., see Olivot *et al.* (2009)) or estimate the amount of tissue that has recovered (e.g., Inoue *et al.* (2014)). These maps can be simply obtained by replacing the contrast parameter value of each voxel by its closest non-inferior threshold:

```
R> calcThresholdMRIaggr(MRIaggr.Pat1_red, param = c("TTP_t0", "MTT_t0"),
```

```
+ threshold = 1:10, update.object = TRUE,
+ name_newparam = c("TTP.th_t0", "MTT.th_t0"))
```

Step 1 : keep both hemispheres, keep CSF

Step 2 : thresholding *****

```
allocContrast[MRIaggr] : Cartographies "TTP.th_t0" "MTT.th_t0"
                        have been allocated
```

```
R> head(selectContrast(MRIaggr.Pat1_red, param = c("TTP_t0", "TTP.th_t0")))
```

	TTP_t0	TTP.th_t0
432	4.56	4
433	3.94	3
434	3.72	3
435	3.59	3
436	3.66	3
437	4.08	4

However, perfusion maps are often influenced by artefacts and noise that bias the estimation of the extent of the area of interest. For instance, stroke lesion voxels, just as CSF voxels, are characterized by high perfusion delays. By setting argument `rm.CSF` to `TRUE`, the method will search for the CSF, the GM, and the WM parameters and will set the CSF voxels to the lowest threshold. Similarly, voxels from the non-lesioned hemisphere can be set to the lowest threshold by setting argument `hemisphere` to `"lesion"`:

```
R> calcThresholdMRIaggr(MRIaggr.Pat1_red, param = c("TTP_t0", "MTT_t0"),
+ threshold = 1:10, rm.CSF = TRUE, hemisphere = "lesion",
+ update.object = TRUE, name_newparam = c("TTP.red_t0", "MTT.red_t0"))
```

Step 1 : keep only the "lesion" hemisphere, remove CSF

Step 2 : thresholding *****

```
allocContrast[MRIaggr] : Cartographies "TTP.red_t0" "MTT.red_t0"
                        have been allocated
```

Finally a "growing region" step ([Revol and Jourlin \(1997\)](#)) can be added to remove more artefacts. Given a threshold τ , this method generates a region from initial seeds by including the neighboring voxels whose contrast values are above τ . Voxels inside this region are given value τ and voxels outside it keep their previous values. The user has to specify the neighboring range (argument `W.range`) and the initial seeds (argument `seed`). Lesioned voxels can be used to define the initial seeds:

```
R> calcThresholdMRIaggr(MRIaggr.Pat1_red, param = c("TTP_t0", "MTT_t0"),
+ threshold = 1:10, rm.CSF = TRUE, hemisphere = "lesion",
+ GRalgo = TRUE, W.range = sqrt(2), as.logical = TRUE,
+ seed = c("MASK_T2_FLAIR_t2", "MASK_DWI_t0"),
+ update.object = TRUE, name_newparam = c("TTP.GR_t0", "MTT.GR_t0"))
```

computing W ...

Step 1 : keep only the "lesion" hemisphere, remove CSF

Step 2 : thresholding *****

Step 3 : Growing Region

```
TTP_t0 *****
```

MTT_t0 *****

```
allocContrast[MRIfaggr] : Cartographies "TTP.GR_t0" "MTT.GR_t0"
                           have been allocated
```

Here, the lesioned voxels from the admission and follow-up periods were used to initialize the growing region with the lowest threshold, then the region obtained with the lower threshold is used to initialize the algorithm of the following threshold. The raw parameter and the result of each thresholding operation are displayed on Figure 6.

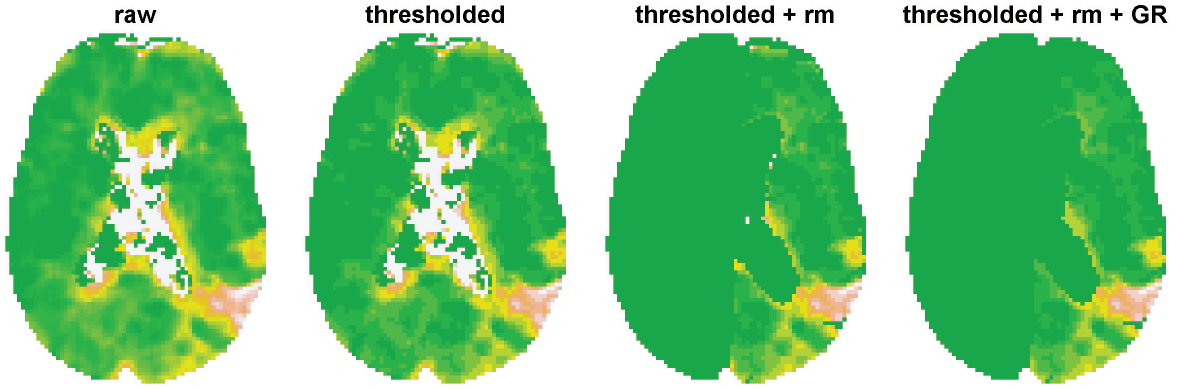


Figure 6: Raw TTP parameter map, thresholded parameter map, thresholded parameter map after the removal of the contralateral hemisphere and the CSF-related voxels, and thresholded parameter map after application of a growing region algorithm in addition to the removal of the contralateral hemisphere and the CSF-related voxels.

4.4. Contrast parameter normalization

In order to normalize contrast parameter values between patients, two methods are proposed: `calcContralateral` that normalizes each voxel with respect to its contralateral correspondent and `calcNormalization` that normalizes parameter values with respect to specific brain regions.

For each voxel, `calcContralateral` computes the difference between its contrast and its contralateral voxel contrast. The contralateral voxel contrast can be estimated in two ways: i) taking the mean (or the median) contrast of the contralateral voxels that overlap the projection of the voxel in the contralateral hemisphere; ii) associating each voxel with the contralateral voxel that optimizes a compromise between contrast similarity and spatial proximity using a single contrast parameter as reference:

$$contralateral_i = \underset{j \in \text{contralateral}}{\operatorname{argmin}} \frac{|Xref(i) - Xref(j)|}{\sigma} + \lambda \operatorname{dist}(i, j)$$

$\operatorname{dist}(i, j) < D$

where $Xref(i)$ is the contrast at voxel i of the reference parameter
 $contralateral$ is the set of contralateral voxels
 σ is the standard deviation of the contrast
 λ, D are pre-defined parameters (by default $\lambda = 1$ and $D = 1$)
 $dist(i, j)$ is the Euclidean distance between the center of voxel j
 and the projection of voxel i in the contralateral hemisphere

The normalized values of voxels with no contralateral correspondents were set to NA. In practice, the search is limited to a certain range fixed by the D parameter. Argument `type` sets the computation ("mean", "median", or "1NN_penalised") and argument `param.ref` gives the reference contrast parameter.

```
R> calcContralateral(MRIaggr.Pat1_red, param = c("DWI_t0", "TTP_t0"),
+   type = "1NN_penalised", param.ref = "T1_t0",
+   distband = 1, lambda = 1, trace = TRUE, update.object = TRUE)

Left slice : 1 2 3
Right slice : 1 2 3
allocContrast[MRIaggr] : Cartographies "DWI_t0_contro" "TTP_t0_contro"
                        have been allocated
```

Figure 7 displays the results of this normalization for the TTP parameter.

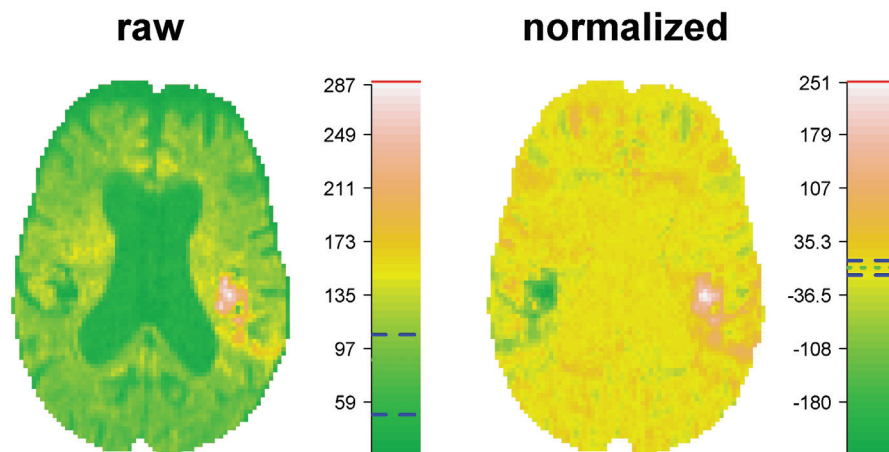


Figure 7: Initial TTP contrast maps (left panel) and normalized TTP contrast map using the contralateral hemisphere (right panel).

Another common normalization is to take the intact WM as reference. For this, `calcNormalization` computes the mean and the standard deviation of the contrast parameter in the contralateral hemisphere and stores them into the `normalization` attribute. Arguments `rm.CSF` and `rm.GM` are set to `TRUE` to remove CSF- and GM-related voxels:

```
R> calcNormalization(MRIaggr.Pat1_red, param = c("DWI_t0", "TTP_t0"),
+   rm.CSF = TRUE, rm.GM = TRUE, update.object = TRUE)

global
slice : 1 2 3
allocNormalization[MRIaggr] : @normalization has been allocated
```


Then the normalized parameters can be extracted with the `selectContrast` method specifying the type of normalization. Argument `norm_mu` indicates the type of centering and argument `norm_sigma` the type of scaling.

```
R> carto <- selectContrast(MRIaggr.Pat1_red, param = "DWI_t0", coords = TRUE)
R> cartoN <- selectContrast(MRIaggr.Pat1_red, param = "DWI_t0",
+   norm_mu = "contralateral", norm_sigma = "contralateral")
R> head(data.frame(carto, DWI_t0_normalized = cartoN$DWI_t0))
```

	i	j	k	DWI_t0	DWI_t0_normalized
432	42	6	1	82.90	-1.24974
433	43	6	1	76.10	-1.70359
434	44	6	1	81.07	-1.37188
435	45	6	1	94.16	-0.49823
436	46	6	1	89.27	-0.82460
437	47	6	1	82.07	-1.30514

The available types of normalization include "contralateral" where the normalization value is computed on the whole contralateral hemisphere, "c_contralateral" where for each slice the normalization value is computed on the contralateral part of the slice and "c3_contralateral" where for each slice the normalization value is computed on the contralateral part of the current slice, the slice above and the slice below.

5. Display tools

5.1. Slice by slice display of 3D images

`multiplot` is the main method to display `MRIdaggr` objects. It performs a slice by slice display of the contrast parameter specified by argument `param`:

```
R> multiplot(MRIaggr.Pat1_red, param = "DWI_t0")
```

See Figure 1 (bottom left panel) to visualize the result. By default, all slices are displayed on one or several graphical devices, each graphical device containing up to 9 slices. Argument `num` can be used to specify the slices that should be displayed: for instance `num=1:2` will display only the first two slices. A color legend is automatically displayed after the last slice. In this legend, the quantiles are represented by lines (red lines for the minimum and maximum, blue lines for the first and third quantiles and a green line for the median). Quantiles can be removed by setting argument `quantiles.legend` to `FALSE` and the legend will not be displayed if `legend` equals `FALSE`. By default, the title plot of each slice contains the patient identifier, the name of the contrast parameter, and the slice number. The slice number can be removed by setting `num.main` to `FALSE` and the rest of the title will take the value of the `main` argument if it is set to a non-NULL value. The axes can be also removed by setting `axes` to `FALSE`. For instance, the last panel of Figure 6 can be generated using:

```
R> multiplot(MRIaggr.Pat1_red, param = "TTP.GR_t0", num = 1,
+   window = NULL, legend = FALSE, axes = FALSE,
+   main = "thresholded + rm + GR", num.main = FALSE)
```

Argument `window` specifies how to manage the graphical devices. It can be set to "eps",

"pdf", "png", "svg" (to export each device in, respectively, eps, pdf, png, or svg file) or to NULL (to disable the automatic management). In the latter case, `multiplot` will plot each slice leaving to the user the choice of opening and splitting the graphical devices. For example, it may be interesting to display different contrast parameters on the same graphical device like in Figure 6:

```
R> par(mfrow = c(1, 4), mar = c(0, 1, 2, 1))
R> multiplot(MRIaggr.Pat1_red, param = "TTP_t0", num = 1,
+   breaks = seq(0, 10, 0.01), window = NULL, legend = FALSE,
+   axes = FALSE, main = "raw", num.main = FALSE)
R> multiplot(MRIaggr.Pat1_red, param = "TTP.th_t0", num = 1,
+   window = NULL, legend = FALSE, axes = FALSE,
+   main = "thresholded", num.main = FALSE)
R> multiplot(MRIaggr.Pat1_red, param = "TTP.red_t0", num = 1,
+   window = NULL, legend = FALSE, axes = FALSE,
+   main = "thresholded + rm", num.main = FALSE)
R> multiplot(MRIaggr.Pat1_red, param = "TTP.GR_t0", num = 1,
+   window = NULL, legend = FALSE, axes = FALSE,
+   main = "thresholded + rm + GR", num.main = FALSE)
```

By default, `multiplot` uses the `image` function to display the slices. It extracts the appropriate voxel coordinates and contrasts and transforms them into a matrix format to be readable by the `image` function. The color of each voxel is inferred from a color palette and a scale of contrast values. The default color palette is `"terrain.colors"` but can be changed to other color palettes (e.g., `"grey.colors"`) or to a specific set of colors (e.g., `c("blue", "green", "orange", "red")`) with argument `palette`. Argument `breaks` specifies either the number of categories of the color scale or the contrast values used to group the voxels into categories. In Figure 5, the color intervals varied from 0 to 250 with increments of 1 for the first four panels and varied from -50 to 50 with increments of 1 for the last two panels:

```
R> par(mfrow = c(2, 3), mar = c(0.5, 0.5, 3, 0.5))
R> multiplot(MRIaggr.Pat1_red, param = "DWI_t0", num = 1,
+   breaks = 1:250, window = NULL, legend = FALSE, axes = FALSE,
+   num.main = FALSE, main = "raw")
R> multiplot(MRIaggr.Pat1_red, param = "DWI_t0_3D_G3", num = 1,
+   breaks = 1:250, window = NULL, legend = FALSE, axes = FALSE,
+   num.main = FALSE, main = "Gaussian filter")
R> multiplot(MRIaggr.Pat1_red, param = "DWI_t0_w3D_G3", num = 1,
+   breaks = 1:250, window = NULL, legend = FALSE, axes = FALSE,
+   num.main = FALSE, main = "edge preserving \n Gaussian filter")
R> multiplot(MRIaggr.Pat1_red, param = "DWI_t0_3D_M3", num = 1,
+   breaks = 1:250, window = NULL, legend = FALSE, axes = FALSE,
+   num.main = FALSE, main = "median filter")
R> multiplot(MRIaggr.Pat1_red, param = "DWI_t0_2D_Sx", num = 1,
+   breaks = (-50):50, window = NULL, legend = FALSE, axes = FALSE,
+   num.main = FALSE, main = "Sobel filter (vertical)")
R> multiplot(MRIaggr.Pat1_red, param = "DWI_t0_w2D_Sx", num = 1,
```

```
+ breaks = (-50):50, window = NULL, legend = FALSE, axes = FALSE,
+ num.main = FALSE, main = "edge preserving \n Sobel filter (vertical)")
```

When the colors cannot be inferred from a vector of values, `multiplot` calls the `plot` function and passes it through a vector containing the color with which each voxel should be displayed. This occurs when the user defines the color associated with each voxel through argument `col` or when the user requests a multiparametric display. The multiparametric display works only with two to four contrast parameters, values ranging between 0 and 1, and totaling to 1 for each voxel. It is especially useful for displaying class probability memberships as shown on Figure 4 (right panel). The figure was obtained by giving the names of the membership parameters and selecting a multiparametric palette ("rgb" or "hsv"):

```
R> par(mar = rep(2, 4))
R> multiplot(MRIaggr.Pat1_red, param = c("CSF", "WM", "GM"),
+ window = NULL, num = 1, legend = FALSE, palette = "rgb",
+ axes = FALSE, main = "calcTissueType", num.main = FALSE)
```

`multiplot` allows displaying up to three additional sets of points over the initial contrast map by using arguments `index1`, `index2`, and `index3`. This can be performed in two ways: by giving the coordinates of the points or by calling an existing indicator variable in the object. The last panel of Figure 3 was generated with the second option:

```
R> multiplot(MRIaggr.Pat1, param = "T2_GRE_t0", num = 1,
+ index1 = list(coords = "MASK_DWI_t0", cex = 0.5), as.logical = TRUE,
+ index2 = list(coords = "mask", outline = TRUE, cex = 0.75), window = NULL,
+ legend = FALSE, main = "T2, lesion and mask", num.main = FALSE)
```

For the first set of points, the `multiplot` function extracted the MASK DWI t0 values, converted them into logical values, and displayed only the voxels that corresponded to the TRUE values. Additional arguments can be given in the list like the expansion factor of the points (`cex`), the color (`col`), or the label type (`pch`). In the second set of points, element `outline=TRUE` indicated the display of only the voxels on the edge instead of all the voxels of the group.

The `multiplot` method can be similarly applied to `data.frames`. The main difference is that the first two arguments must refer to the voxel coordinates (argument `object`) and to the contrast values (argument `contrast`). For instance, the third panel of Figure 3 can be obtained using:

```
R> multiplot(selectCoords(MRIaggr.Pat1),
+ contrast = res4_a3[, "mask"], num = 3, legend = FALSE,
+ main = "k-means smoothed", num.main = FALSE, window = NULL)
```

5.2. Visual segmentation

Because automatic segmentation may fail, visual segmentation may be useful to perform or correct segmentation. The `outlineMRIaggr` method enables to define a new segmentation by clicking on the graphical device. Each click will sequentially define the outline of the segmentation mask. The segmentation ends when the user clicks on the initial point. Then, the method will automatically fill the outline (except if `fill` is set to `FALSE`) and will return a binary parameter that indicates whether each voxel belongs to the segmented region or not:

```
R> outlineMRIaggr(MRIaggr.Pat1_red, param = "T2_FLAIR_t2", num = 1,
+ update.object = TRUE, name_newparam = "VisualSeg")
```

allocContrast[MRIaggr] : Cartography "VisualSeg" has been allocated

The segmentation process is displayed on Figure 8.

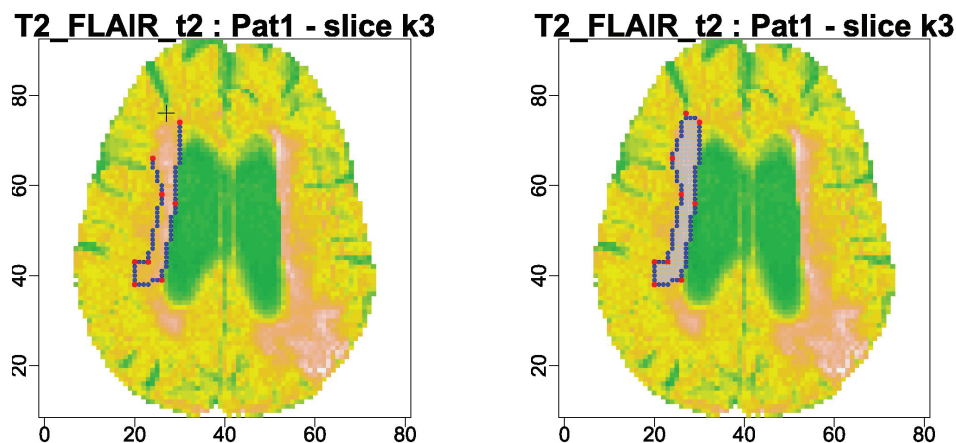


Figure 8: Manual outline of the white matter intensity (left panel: outline in progress; right panel: complete outline). The red points correspond to clicks of the user. The blue points are interpolated points and the grey points are the interior points of the segmented form.

The correction of an existing segmentation can be done by indicating the initial segmentation to argument `index1` and the operator to apply to the result of the visual segmentation (argument `operator_index1`):

```
R> outlineMRIaggr(MRIaggr.Pat1_red, param = "DWI_t0", num = 1,
+ index1 = "MASK_DWI_t0", as.logical = TRUE, operator_index1 = "intersection",
+ update.object = TRUE, name_newparam = "MASK_DWI_t0_corrected")
```

allocContrast[MRIaggr] : Cartography "MASK_DWI_t0_corrected" has been allocated

Because `operator_index1` was set to "intersection" the method retained the intersection between the user segmentation and the first lesion segmentation defined by MASK_DWI_t0 (Figure 9).

5.3. Automatic generation of a pdf report

Visual checking is often the best way to ensure the validity of the processing step or the fit of the model. However, the number of files to examine grows quickly with the number of patients and the contrast parameters to display. Generating a single file ordered by patient and by parameter may ease this checking. The `constLatex` function can be used to automatize the generation of this document. It takes, as argument, the path to a root directory. This directory contains subdirectories which contain the image files. For instance, creating a root directory named "Display" in the current working directory:

```
R> dir.create("Display")
```

Then creating two subdirectories, one per contrast parameter:

```
R> dir.create(file.path("Display", "T1_t0"))
```

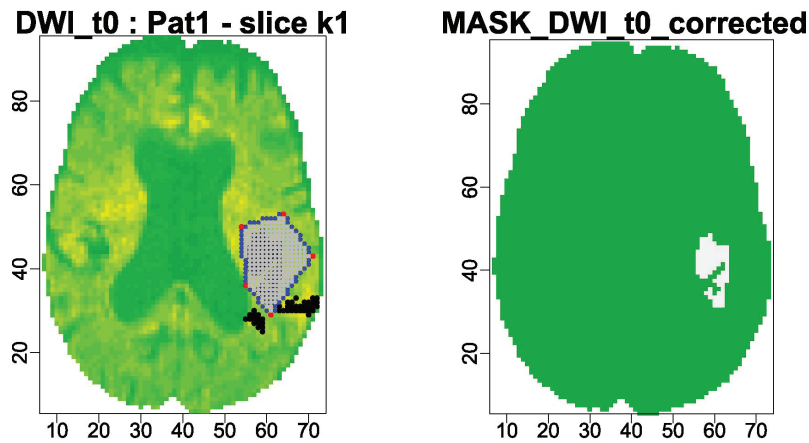


Figure 9: Manual correction of the lesion mask. Left panel: The black points correspond to the lesion mask and the grey points to the user segmentation. Right panel: The white voxels correspond to the corrected mask.

```
R> dir.create(file.path("Display", "T2_GRE_t0"))
```

In each subdirectory, the image filenames should meet specific rules:

- begin with the patient identifier followed by an underscore.
- contain the first and the last slice numbers, separated by a hyphen after the keyword "slice".
- the rest of the filename should be the same for all image files in a subdirectory.

The files exported using the `multiplot` function meet automatically these rules:

```
R> multiplot(MRIaggr.Pat1, param = "T1_t0",
+ mfrow = c(2, 4), axes = FALSE, mar = c(0, 0, 1.5, 0),
+ window = "png", path = "Display/T1_t0/")
R> multiplot(MRIaggr.Pat2, param = "T1_t0",
+ mfrow = c(3, 6), axes = FALSE, mar = c(0, 0, 1.5, 0), legend = FALSE,
+ window = "png", path = "Display/T1_t0/")
R> multiplot(MRIaggr.Pat1, param = "T2_GRE_t0",
+ mfrow = c(2, 4), axes = FALSE, mar = c(0, 0, 1.5, 0),
+ window = "png", path = "Display/T2_GRE_t0/")
R> multiplot(MRIaggr.Pat3, param = "T2_GRE_t0",
+ mfrow = c(2, 4), axes = FALSE, mar = c(0, 0, 1.5, 0),
+ window = "png", path = "Display/T2_GRE_t0/")
R> list.files("Display/T2_GRE_t0/")

[1] "Pat1_multiplot_T2_GRE_t0(kslice1-3_both).png"
[2] "Pat3_multiplot_T2_GRE_t0(kslice1-8_both).png"
[3] "Pat3_multiplot_T2_GRE_t0(kslice17-21_both).png"
[4] "Pat3_multiplot_T2_GRE_t0(kslice9-16_both).png"
```

Then the \LaTeX document can be created by indicating the root directory and the name of the \LaTeX file to create:

```
R> res <- constLatex(dir = "Display", file = "LatexDisplay")

%% 2 directories found in the root directory
%% 3 identifiers found in the subdirectories
% 3 identifiers will be considered according to 'param' argument

%%% num : subsection | directory
% (label)

%%% 1   : T1 t0      | /T1_t0
% label : "T1 t0"

%%% 2   : T2 GRE t0  | /T2_GRE_t0
% label : "T2 GRE t0"

% graph of patient 1(Pat1) 2(Pat2) 3(Pat3)
```

Then corresponding pdf document can be generated using:

```
R> require(tools)
R> texi2pdf(file = "LatexDisplay.tex")
```

A screenshot of the pdf file is displayed on Figure 10. Arguments `constLatex` enable specifying the name of the subsections (argument `subsection`), the order in which the parameters should be displayed (argument `param`), or the legend to display below the figures (argument `label`).

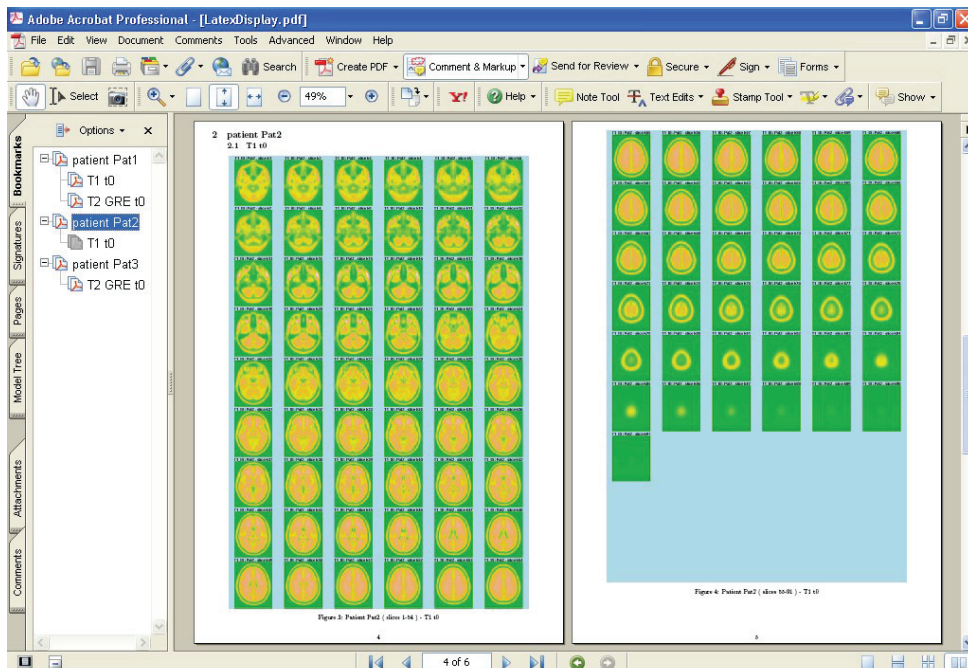


Figure 10: Screenshot of pages 5 and 6 of the \LaTeX document.

6. Conclusion

Package **MRIfaggr** provides a convenient framework for the management of multiparametric medical images. Only the main functionalities of the package are presented here. For a comprehensive presentation of each function, the reader is advised to refer to the documentation on the package. In addition, three demonstration files are present in the package to illustrate further the package functionalities.

Though the package is illustrated here with data on only one patient, it has been already successfully used to process data on 63 stroke patients from the I-know cohort (Østergaard *et al.* (2006)). Future works will design tools for automatic lesion segmentation and an interface with **RNiftyReg** to facilitate image registration.

References

- Arya S, Mount D, Kemp SE, Jefferis G (2015). *RANN: Fast Nearest Neighbour Search (Wraps Arya and Mount's ANN Library)*. R package version 2.5, URL <http://CRAN.R-project.org/package=RANN>.
- Bankman IN, Rogowska J, Sutton MA, Wismueller A, McInerney T, Terzopoulos D, Staib LH, Wang YM, Zeng X, Duncan JS, Xu C, Han X, Prince JL, Atkins MS, Mackiewicz BT (2009). *Handbook of Medical Image Processing and Analysis (Chapters 5 to 11)*. Second edition. Academic Press, Burlington.
- Chambers J (2008). *Software for Data Analysis: Programming with R*. Springer-Verlag.
- Clayden J (2014). *mmmand: Mathematical Morphology in Any Number of Dimensions*. R package version 1.1.0, URL <http://CRAN.R-project.org/package=mmmand>.
- Clayden JD, Maniega SM, Storkey AJ, King MD, Bastin ME, Clark CA (2011). “TractoR: Magnetic Resonance Imaging and Tractography with R.” *Journal of Statistical Software*, **44**(8), 1–18. URL <http://www.jstatsoft.org/v44/i08/>.
- Feng D, Tierney L (2011). “mrirc: A Package for MRI Tissue Classification.” *Journal of Statistical Software*, **44**(7), 1–20. URL <http://www.jstatsoft.org/v44/i07/>.
- Feng D, Tierney L, Magnotta V (2012). “MRI Tissue Classification Using High-Resolution Bayesian Hidden Markov Normal Mixture Models.” *Journal of the American Statistical Association*, **107**(497), 102–119.
- Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, Essen DCV, Jenkinson M (2013). “The Minimal Preprocessing Pipelines for the Human Connectome Project.” *NeuroImage*, **80**(0), 105–124.
- Inoue M, Mlynash M, Christensen S, Wheeler HM, Straka M, Tipirneni A, Kemp SM, Zaharchuk G, Olivot JM, Bammer R, Lansberg MG, Albers GW (2014). “Early Diffusion-Weighted Imaging Reversal After Endovascular Reperfusion Is Typically Transient in Patients Imaged 3 to 6 Hours After Onset.” *Stroke*, **45**(4), 1024–1028.

- Kurhanewicza J, Vignerona D, Carrollb P, Coakleya F (2008). “Multiparametric Magnetic Resonance Imaging in Prostate Cancer: Present and Future.” *Current Opinion in Urology*, **18**(1), 71–7.
- Muir KW, Buchan A, von Kummer R, Rother J, Baron JC (2006). “Imaging of Acute Stroke.” *The Lancet Neurology*, **5**(9), 755–768.
- Olivot JM, Mlynash M, Thijs VN, Kemp S, Lansberg MG, Wechsler L, Bammer R, Marks MP, Albers GW (2009). “Optimal Tmax Threshold for Predicting Penumbra Tissue in Acute Stroke.” *Stroke*, **40**(2), 469–475.
- Østergaard L, Institut National de la Santé et de Recherche Médicale and Université Claude Bernard, Fundació Privada Institut d'Investigació Biomèdica de Girona, University of Cambridge, Universitätsklinikum Hamburg-Eppendorf, Systematic Software Engineering and Dimac (2006). “I-know Integration Information from Molecule to Man.” *European Commission's Sixth Framework Programme*. URL http://vph-portal.eu/documents/10180/35174/I_Know_factsheet.pdf.
- Pau G, Fuchs F, Sklyar O, Boutros M, Huber W (2010). “EBImage - An R Package for Image Processing with Applications to Cellular Phenotypes.” *Bioinformatics*, **26**(7), 979–981. doi:10.1093/bioinformatics/btq046.
- Reich DS, Smith SA, Zackowski KM, Gordon-Lipkin EM, Jones CK, Farrell JAD, Mori S, van Zijl PCM, Calabresi PA (2007). “Multiparametric Magnetic Resonance Imaging Analysis of the Corticospinal Tract in Multiple Sclerosis.” *Neuroimage*, **38**(2), 271–279.
- Revol C, Jourlin M (1997). “A New Minimum Variance Region Growing Algorithm for Image Segmentation.” *Pattern Recognition Letters*, **18**(3), 249 – 258.
- Tabelow K, Polzehl J (2011). “Statistical Parametric Maps for Functional MRI Experiments in R : The Package fmri.” *Journal of Statistical Software*, **44**(1-21), 11. URL <http://www.jstatsoft.org/v44/i11/>.
- Tabelow K, Whitcher B (2011). “Special Volume on Magnetic Resonance Imaging in R.” *Journal of Statistical Software*, **44**(1).
- Tomasi C, Manduchi R (1998). “Bilateral Filtering for Gray and Color Images.” *Computer Vision, 1998. Sixth International Conference on*, pp. 839–846. doi:10.1109/ICCV.1998.710815.
- Whitcher B, Schmid VJ, Thornton A (2011). “Working with the DICOM and NIfTI Data Standards in R.” *Journal of Statistical Software*, **44**(6), 1–28. URL <http://www.jstatsoft.org/v44/i06/>.
- Wu O, Dijkhuizen RM, Sorensen AG (2010). “Multiparametric MR Imaging of Brain Disorders.” *Topics in Magnetic Resonance Imaging*, **21**(2), 129–138.

Affiliation:

Ozenne Brice

Université de Lyon, Département Biomaths-Santé

Université Lyon 1, Département Biomaths-Santé

Hospices Civils de Lyon, Service de Biostatistique

LBBE CNRS UMR 5558, Équipe Biostatistique-Santé

165 Chemin du Grand Revoyet

Pierre-Bénite 69310, France

E-mail: brice.ozenne@chu-lyon.fr

Journal of Statistical Software

published by the American Statistical Association

Volume VV, Issue II

MMMMMM YYYY


<http://www.jstatsoft.org/>

<http://www.amstat.org/>

Submitted: yyyy-mm-dd

Accepted: yyyy-mm-dd

2.5 Nombre de téléchargements de la librairie depuis mars 2015

La librairie *MRIaggr* est librement téléchargeable depuis le CRAN (The Comprehensive R Archive Network). Le CRAN redirige l'utilisateur vers différents serveurs (appelés miroirs) afin que celui-ci effectue le téléchargement depuis un serveur géographiquement proche. Il existe 102 miroirs dont 4 en France (deux à Lyon, un à Paris et un à Montpellier). Seul un miroir du CRAN propose de récupérer le nombre de téléchargements : il s'agit du miroir de Rstudio (<http://cran-logs.rstudio.com/>), Rstudio étant un environnement de développement très populaire pour le langage . Le nombre de téléchargements hebdomadaires de la librairie *MRIaggr* est représenté sur la figure III.20. Il est relativement stable au cours du temps avec un peu plus de 40 téléchargements par semaine pour un total de 880 téléchargements (1130 si l'on compte les téléchargements répétés depuis une même adresse IP). Le package est le plus souvent téléchargé une seule fois (79,5 %), parfois deux fois (17,5 %), le maximum étant de 13 fois. L'origine des téléchargements montre une large couverture internationale du miroir de Rstudio avec principalement l'Asie (41,2%), l'Amérique (29,2%) et l'Europe (26,3%). L'Afrique est en revanche sous-représentée (1,08%).

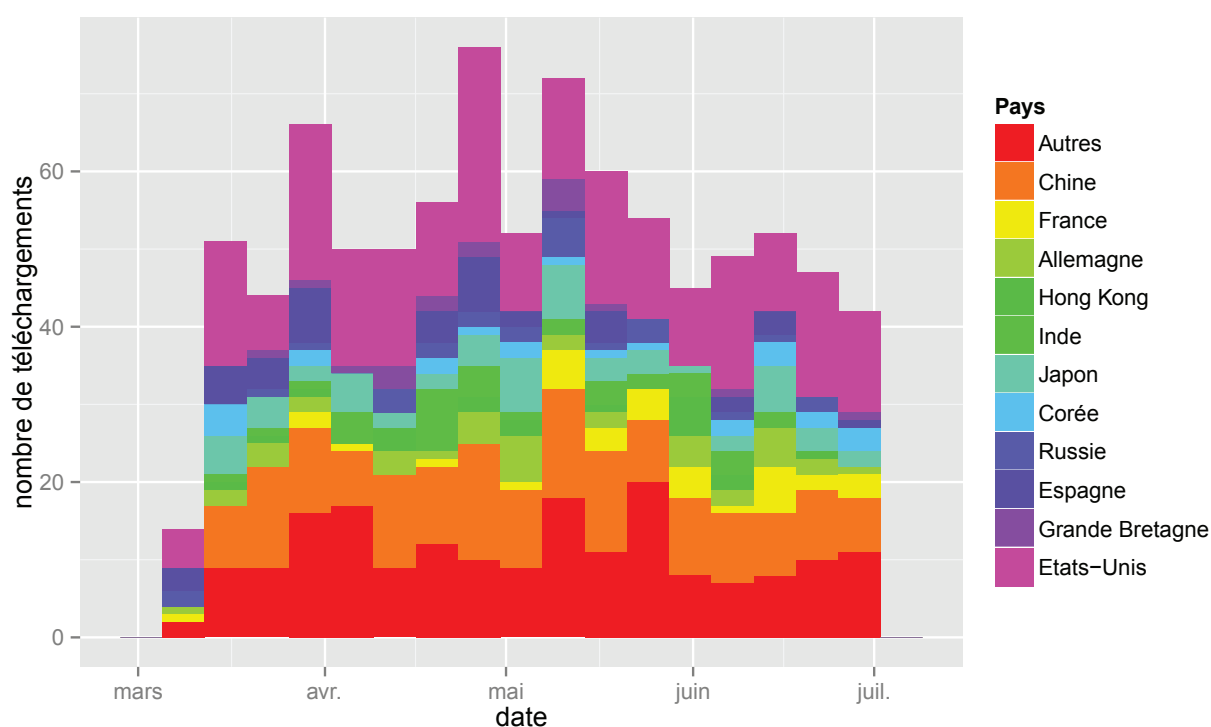





FIGURE III.20 – Nombre de téléchargements hebdomadaires de la librairie *MRIaggr* depuis le miroir de Rstudio entre mars et juillet 2015. Les téléchargements effectués avec une même adresse IP ne comptent que pour 1.

Le nombre de téléchargements de la librairie *MRIaggr* est un peu inférieur à celui des librairies présentées dans le tableau (III.2). La librairie *oro.nifti* est la plus téléchargée avec 3131 téléchargements, ce qui est attendu puisque de nombreuses librairies dépendent de *oro.nifti*.

Pour les autres librairies le nombre de téléchargements variait entre 1303 et 1736, ce qui reste du même ordre de grandeur que les 1130 téléchargements enregistrés pour la librairie que nous proposons.

2.6 Perspectives

La librairie *MRIfaggr* propose de nombreux outils pour la gestion et le pré-traitement de données d'imagerie médicale, dont certains sont particulièrement adaptés à l'imagerie cérébrale. La vocation de cette librairie est de proposer un cadre de travail général pour l'analyse de données d'imagerie sous  et non de se substituer aux logiciels spécialisés évoqués dans la section 2.1. Des fonctions ont donc été prévues pour faciliter le transfert de données avec des librairies externes.

L'absence de fonction de segmentation de lésion est une lacune pour la librairie et, plus généralement, pour le logiciel . En effet, bien souvent en imagerie médicale, l'analyse nécessite d'identifier une lésion ou une zone de contraste anormal. Actuellement, sous , trois options sont possibles. La librairie *dpmixsim* propose une méthode de segmentation par modèle de mélange univarié avec estimation automatique du nombre de groupes. Cependant, l'absence de prise en compte de l'information spatiale est limitante : dans notre cas cette méthode ne fournissait pas des résultats satisfaisants. La librairie *EBImage* propose aussi des outils de segmentation mais peu transférables à l'analyse de données cérébrales. Enfin, le modèle de mélange de la librairie *mrirc* est analogue au modèle de mélange avec régularisation locale que nous avons développé. Cependant il est sensible aux artéfacts présents sur les coupes basses ou causés par la maladie de la substance blanche et n'est implémenté que pour une caractérisation univariée de la lésion. L'algorithme de segmentation que nous avons développé a donc toute sa place dans notre librairie. Nous travaillons actuellement à l'optimisation du code et à l'automatisation de la phase d'initialisation. L'objectif est que dans une version finale l'utilisateur doive uniquement spécifier :

- le(s) paramètre(s) de contraste utilisé(s) pour identifier la lésion (typiquement DWI ou T2 FLAIR).
- le paramètre de contraste utilisé pour l'initialisation (typiquement le DWI).
- la portée du voisinage local et celle du voisinage régional.
- les variables à considérer dans le modèle de structure.

Dans un second temps, nous souhaitons intégrer l'algorithme de lissage d'images proposé par l'article de STORATH et al., 2014. Il peut en effet avoir plusieurs applications :

- outil de lissage des paramètres IRM pour éliminer le bruit tout en conservant un bon contraste
- intégration dans l'algorithme de segmentation pour remplacer la phase de régularisation locale. Il pourrait fournir des lésions avec une meilleure cohérence spatiale par rapport à ce qui est actuellement implémenté.

D'autres aspects de la librairie restent à améliorer : la gestion mémoire n'est pas optimisée, ce qui limite la portée de la librairie pour des images à haute résolution. L'utilisation de formats spécialisés proposés (librairies *data.table* et *bigmemory*) pourrait être une voie de solution. Par ailleurs, des améliorations sont nécessaires pour permettre l'utilisation de la librairie par un médecin. Un premier pas consisterait à proposer une fonction unique effectuant l'ensemble de traitements et fournissant en sortie les prédictions volumiques ainsi que les cartographies de risque dans le cas avec et sans reperfusion (figure III.21).

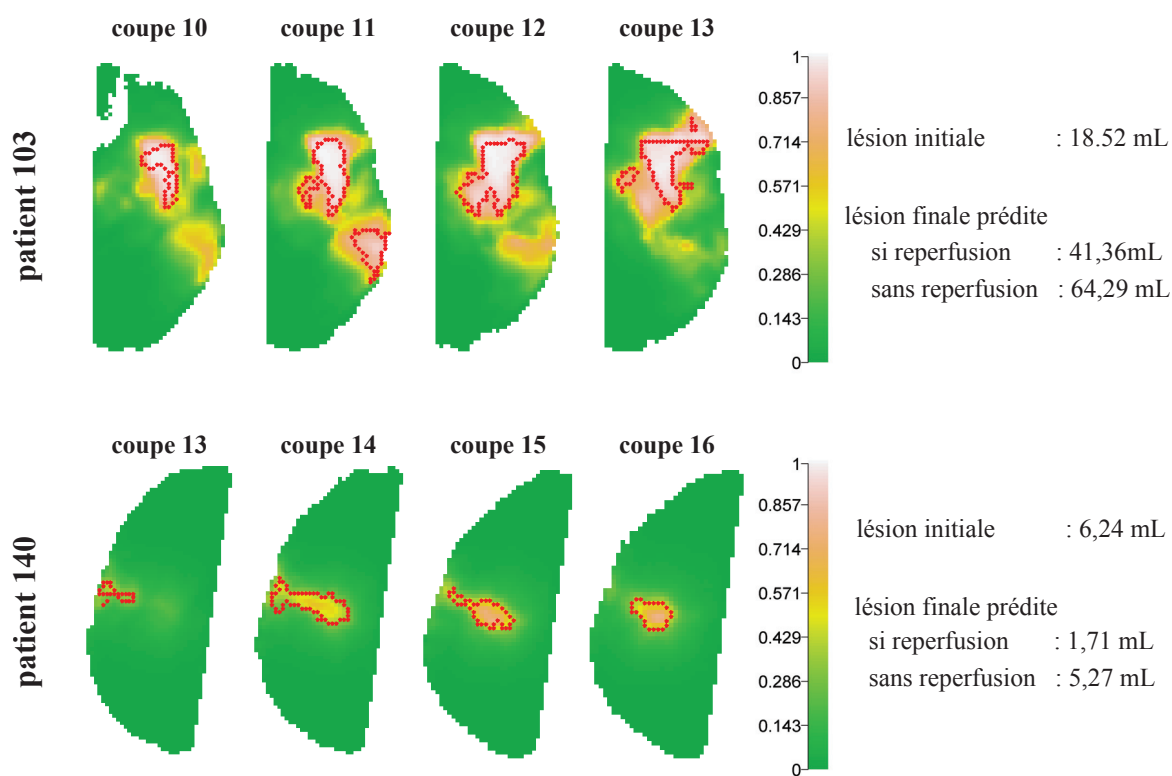


FIGURE III.21 – Exemple de sortie logiciel pouvant aider le médecin dans sa prise de décision. Chaque image représente la probabilité de nécrose prédite entre 0 (en vert) et 1 (en blanc). La lésion en diffusion est contournée en rouge. Premier cas : patient avec une zone à risque plus étendue que la lésion initiale. Second cas : patient avec une zone à risque plus petite que la lésion initiale. Pour ce patient, la lésion en diffusion surestimait la lésion ce qui est confirmé à l'IRM finale puisque le volume de lésion final est de 2,40 mL.

3 Conclusion

Dans cette seconde partie, nous avons développé et implémenté les outils de traitement d'images nécessaires à l'utilisation de modèles prédictifs. Ces outils sont rassemblés dans la librairie de fonctions *MRImgr*, librement téléchargeable sur le site du CRAN. Nous assurons ainsi la reproductibilité des résultats obtenus et facilitons la réutilisation des développements proposés dans cette thèse pour de futurs travaux. Même si le nombre de téléchargements est un indicateur très imparfait, il semble tout de même que notre librairie ait trouvé rapidement un public. L'article soumis au Journal of Statistical Software devrait renforcer sa visibilité et faciliter son utilisation.

Il nous reste cependant à intégrer dans cette librairie l'outil de segmentation de lésion que nous avons développé. Celui-ci comble un manque puisqu'il n'existait pas d'outil adapté à la séquence T2 FLAIR, séquence de référence pour l'identification de l'AVC final. Nous avons donc développé et validé sur données simulées et sur données réelles notre algorithme de segmentation. Celui-ci permet d'effectuer une segmentation multiparamétrique de la lésion intégrant une étape de contrôle de la forme de la lésion afin d'être robuste aux artéfacts d'imagerie. En particulier, nous avons montré que notre algorithme permettait de limiter les erreurs de segmentation causées par la maladie de la substance blanche. La mise à disposition de cet outil permet d'éviter une étape de correction fastidieuse pouvant ainsi faire gagner un temps précieux.


Conclusion générale

Les progrès techniques en imagerie médicale ont été considérables au cours des vingt dernières années : amélioration de la résolution spatiale, de la rapidité d'acquisition et définition de nouveaux paramètres de contraste. Concernant l'AVC, ces progrès techniques peinent cependant à se traduire en bénéfice clinique. Malgré de nombreuses études, aucun paramètre d'imagerie ne s'est révélé suffisamment fiable pour identifier les patients susceptibles de répondre au traitement. Les travaux de recherche, initialement situés en traitement du signal et en analyse d'images, se sont donc tournés vers la biostatistique afin de mieux exploiter l'information disponible. Nous avons donc mis en place les outils nécessaires à l'utilisation de modèle prédictifs pour l'AVC. Ces modèles ont un double intérêt : prédire le bénéfice du traitement en modélisant deux évolutions possibles pour le patient (avec ou sans traitement), et permettre une meilleure quantification de l'effet traitement.

Notre contribution s'est effectuée au niveau méthodologique :

- **en proposant une meilleure évaluation des modèles prédictifs** : nous avons montré que l'indice classiquement utilisé, l'AUC, donnait une vision trop optimiste des performances des modèles contrairement à l'AUPRC qui semblait bien adaptée. Nous avons donc suggéré d'évaluer le modèle sur 3 critères : l'AUPRC, l'AUC (qui garde un intérêt pour la comparaison d'études) et l'erreur volumique. Cette dernière a toute son importance puisque les volumes prédits par le modèle sont vus comme des outils d'aide à la décision. Dans nos études, ces trois indices donnaient des résultats concordants : ils permettaient une évaluation plus précise tout en étant facilement interprétable.
- **en déterminant un modèle prédictif optimal** : nous avons comparé différentes approches prédictives vis-à-vis de la modélisation utilisée, du traitement des données d'imagerie et du choix des variables prédictives. Nous avons montré la supériorité des approches multivariées par rapport aux approches univariées ainsi qu'un léger gain après lissage des données d'imagerie. Nous avons montré l'importance du choix des variables prédictives puisque l'inclusion de variables spatiales améliorait notablement les performances prédictives. Pour ce qui est de la modélisation, seule une meilleure prise en compte de la variabilité inter-patients apparaît prometteuse.
- **en fournissant un cadre statistique à la prédiction volumique** : en reformulant le problème de prédiction volumique au sein d'un modèle de régression fonctionnel. Cela permet de disposer des outils d'inférence usuels pour évaluer l'influence du traitement et permet d'intégrer l'information clinique du patient dans la prédiction.

Ainsi qu'au niveau pratique :

- **en développant un algorithme de segmentation de lésion** : adapté aux difficultés liées aux séquences d'imagerie pour l'AVC. L'algorithme proposé nécessite uniquement d'initialiser la position de la lésion et permet une caractérisation multivariée de la lésion. Il contient deux étapes de correction de forme de la lésion, notamment une généralisation du modèle de Potts, qui améliorent la robustesse de la segmentation en présence de maladie de la substance blanche.
- **en mettant à disposition de la communauté les outils développés** : à travers la librairie de fonctions *MRImaggr* du logiciel . Elle assure à la fois la reproductibilité des résultats obtenus mais aussi, grâce aux différents outils de gestion de données proposés, facilite de futurs travaux. L'algorithme de segmentation, actuellement disponible sur demande, sera prochainement intégré dans la librairie.

La principale limite de notre travail vient du manque de précision du modèle volumique. En effet même si nous avons amélioré la précision volumique des prédictions, elles restent toujours trop imprécises pour être utilisées en clinique. Nous avons néanmoins proposé, en preuve de concept, des outils d'aide à la décision et des critères d'éligibilité au traitement basés sur ces prédictions. Nous espérons que la méthodologie statistique développée dans cette thèse permette une meilleure intégration et évaluation des futurs progrès en imagerie médicale et en modélisation statistique. Enfin, en donnant un cadre statistique à l'évaluation de l'effet traitement au sein du modèle prédictif, nous espérons contribuer à une meilleure identification des traitements prometteurs en phase précoce.

Production scientifique

Publications

Acceptées :

Ozenne B., Subtil F., Østergaard L., and Maucort-Boulch D. *Spatially regularized mixture model for lesion segmentation with application to stroke patients*. Biostatistics (2015) 16(3) :580-595.

Ozenne B., Subtil F., and Maucort-Boulch D. *The precision-recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases*. Journal of clinical epidemiology (2015) 68(8) :855-859.

Ozenne B.*, Cho T.H.*, Mikkelsen I.K., Hermier M., Ribe L., Thomalla G., Pedraza S., Baron J.C., Roy P., Berthezène Y., Nighoghossian N., Østergaard L., and Maucort-Boulch D. *Evaluation of early reperfusion criteria in acute ischemic stroke*. Journal of NeuroImaging (2015) *Publié en ligne*. (*contributions égales)

Hermitte L., Cho T.H., **Ozenne B.**, Nighoghossian N., Mikkelsen I.K., Ribe L., Baron J.C., Østergaard L., Derex L., Hjort N., Fiehler J., Pedraza S., Hermier M., Maucort-Boulch D., and Berthezène Y. *Very low cerebral blood volume predicts parenchymal hematoma in acute ischemic stroke*. Stroke (2013) 44(8) :2318-20.

Soumises :

Ozenne B. and Maucort-Boulch D. *MRlAggr : An R package for multiparametric medical images*. Journal of statistical software. *Prescreening stage*.

En cours :


Ozenne B., Cho T.H., Mikkelsen I.K., Hermier M., Ribe L., Thomalla G., Pedraza S., Baron J.C., Roy P., Berthezène Y., Nighoghossian N., Østergaard L., and Maucort-Boulch D. *Individualized Quantification of the Benefit from Reperfusion Therapies using Stroke Predictive Models*. Stroke. *Finalisation pour soumission*.

Péron J., Buyse M., **Ozenne B.**, Roche L., and Roy P. *An extension to generalized pairwise comparisons for prioritized outcomes with censoring*. Clinical Trials. *En cours de re-soumission*.

Tozlu C., **Ozenne B.**, Cho T.H., Mikkelsen I.K., Hermier M., Ribe L., Thomalla G., Pedraza S., Baron J.C., Roy P., Berthezène Y., Nighoghossian N., and Maucort Boulch D. *Comparison of classification methods for the identification of tissue outcome after ischemic stroke. En cours d'écriture.*

Logiciel

Développement et maintenance de la librairie  **MRIdaggr** portant sur la gestion et le traitement de données IRM multivariées.

Développement de la librairie  **BuyseTest** implémentant un test de comparaison par paires généralisée. Le package a été développé en collaboration avec Julien Péron qui a effectué les tests de validation. Cette librairie sera prochainement soumise au CRAN.

Communications Orales

Ozenne B. and Maucort-Boulch D. *MRIdaggr : un package pour la gestion et le traitement de données multivariées d'imagerie.* 4^e rencontres R, 2015, 24 - 26 juin, Grenoble, France.

Ozenne B., Subtil F., and Maucort-Boulch D. *Lesion Segmentation using a Spatially Regularized Mixture Model.* Applied Statistics 2014, 21 - 24 septembre, Ribno (Bled), Slovenie.

Communications affichées

Ozenne B., Subtil F., and Maucort-Boulch D. *Segmentation de lésion par modèle de mélange avec régularisation spatiale.* EPICLIN 2014, 14-16 mai, Bordeaux, France.

Ozenne B., Subtil F., and Maucort-Boulch D. *Improving lesion segmentation with spatial regularization.* IBC 2014, 6 - 11 juillet, Florence, Italie.

Références

- ALBERS, G. W., V. N. THIJS, L. WECHSLER, S. KEMP, G. SCHLAUG, E. SKALABRIN, R. BAMMER, W. KAKUDA, M. G. LANSBERG, A. SHUAIB, W. COPLIN, S. HAMILTON, M. MOSELEY et M. P. MARKS (2006). « Magnetic resonance imaging profiles predict clinical response to early reperfusion : The diffusion and perfusion imaging evaluation for understanding stroke evolution (DEFUSE) study ». Dans : *Annals of Neurology* 60.5, p. 508–517. URL : <http://dx.doi.org/10.1002/ana.20976>.
- ALLEN, C. L. et U. BAYRAKTUTAN (2008). « Risk factors for ischaemic stroke ». Dans : *International Journal of Stroke* 3.2, p. 105–116. URL : <http://dx.doi.org/10.1111/j.1747-4949.2008.00187.x>.
- ARAKAWA, S., P. M. WRIGHT, M. KOGA, T. G. PHAN, D. C. REUTENS, I. LIM, M. R. GUNAWAN, H. MA, N. PERERA, J. LY, J. ZAVALA, G. FITT et G. A. DONNAN (2006). « Ischemic Thresholds for Gray and White Matter : A Diffusion and Perfusion Magnetic Resonance Study ». Dans : *Stroke* 37.5, p. 1211–1216. URL : <http://dx.doi.org/10.1161/01.STR.0000217258.63925.6b>.
- ARTZI, M., O. AIZENSTEIN, T. JONAS-KIMCHI, V. MYERS, H. HALLEVI et D. B. BASHAT (2013). « FLAIR lesion segmentation : Application in patients with brain tumors and acute ischemic stroke ». Dans : *European Journal of Radiology* 82.9, p. 1512–1518. URL : <http://dx.doi.org/10.1016/j.ejrad.2013.05.029>.
- AUGUSTIN, N. H., M. A. MUGGLESTONE et S. T. BUCKLAND (1996). « An Autologistic Model for the Spatial Distribution of Wildlife ». Dans : *Journal of Applied Ecology* 33.2, pp. 339–347. URL : <http://dx.doi.org/10.2307/2404755>.
- BAGHER-EBADIAN, H., K. JAFARI-KHOUSANI, P. D. MITSIAS, M. LU, H. SOLTANIAN-ZADEH, M. CHOPP et J. R. EWING (2011). « Predicting Final Extent of Ischemic Infarction Using Artificial Neural Network Analysis of Multi-Parametric MRI in Patients with Stroke ». Dans : *PLoS ONE* 6.8, e22626. URL : <http://dx.doi.org/10.1371/journal.pone.0022626>.
- BAMBER, D. (1975). « The area above the ordinal dominance graph and the area below the receiver operating characteristic graph ». Dans : *Journal of Mathematical Psychology* 12.4, p. 387–415. URL : [http://dx.doi.org/10.1016/0022-2496\(75\)90001-2](http://dx.doi.org/10.1016/0022-2496(75)90001-2).
- BARDET, J. (2007). « La prise en charge précoce des accidents vasculaires cérébraux ». Dans : *Office parlementaire d'évaluation des politiques de santé*. URL : <http://www.assemblee-nationale.fr/13/pdf/rap-off/i0236.pdf>.
- BATES, D. M. (2010a). *lme4 : Mixed-effects modeling with R*. Springer. URL : https://r-forge.r-project.org/scm/viewvc.php/*checkout*/www/lmmwR/lrgprt.pdf?revision=600&root=lme4.

- BATES, D. M. (2010b). *Using lme4 : Mixed-effects Modeling in R*. URL : <http://www.stat.wisc.edu/~bates/UseR2008/WorkshopD.pdf>.
- BAUER, S., R. WIEST, L.-P. NOLTE et M. REYES (2013). « A survey of MRI-based medical image analysis for brain tumor studies ». Dans : *Physics in Medicine and Biology* 58.13, R97–R129. URL : <http://dx.doi.org/10.1088/0031-9155/58/13/r97>.
- BEAUCHAMP, N. J. et R. N. BRYAN (1998). « Acute Cerebral Ischemia Infarction : A Pathophysiologic Review and Radiologic Perspective ». Dans : *American Journal of Roentgenology* 171.1, p. 73–84. URL : <http://dx.doi.org/10.2214/ajr.171.1.9648768>.
- BÉJOT, Y, B DAUBAIL, A JACQUIN, J DURIER, G. OSSEBY, O ROUAUD et M. GIROUD (2014). « Trends in the incidence of ischaemic stroke in young adults between 1985 and 2011 : the Dijon Stroke Registry ». Dans : *Journal of Neurology, Neurosurgery, and Psychiatry* 85.5, p. 509–13. URL : <http://dx.doi.org/10.1136/jnnp-2013-306203>.
- BENJAMINI, Y. et D. YEKUTIELI (2001). « The control of the false discovery rate in multiple testing under dependency ». Dans : *The Annals of Statistics* 29.4, p. 1165–1188. URL : http://www.math.tau.ac.il/~yekutieli/papers/dep_annals.pdf.
- BESAG, J. (1974). « Spatial Interaction and the Statistical Analysis of Lattice Systems ». Dans : *Journal of the Royal Statistical Society. Series B (Methodological)* 36.2, p. 192–236. URL : http://dx.doi.org/10.1007/978-1-4612-0667-5_13.
- (1975). « Statistical Analysis of Non-Lattice Data ». Dans : *Journal of the Royal Statistical Society. Series D* 24.3, 179–195. URL : <http://dx.doi.org/10.2307/2987782>.
- BISHOP, C. M. (2007). *Pattern Recognition and Machine Learning*. Computer Science. Springer. URL : <http://research.microsoft.com/en-us/um/people/cmbishop/prml/>.
- BLANCHET, J. (2007). « Modèles markoviens et extensions pour la classification de données complexes ». Dans : *Université Joseph-Fourier - Grenoble I*. URL : <https://hal.inria.fr/tel-00195271v1/document>.
- BOUIX, S., M. MARTIN-FERNANDEZ, L. UNGAR, M. NAKAMURA, M.-S. KOO, R. W. MCCARLEY et M. E. SHENTON (2007). « On evaluating brain tissue classifiers without a ground truth ». Dans : *NeuroImage* 36.4, p. 1207–1224. URL : <http://dx.doi.org/10.1016/j.neuroimage.2007.04.031>.
- BOUTS, M. J. R. J., I. A. C. W. TIEBOSCH1, A. van der TOORN, M. A. VIERGEVER1, O. WU et R. M. DIJKHUIZEN1 (2013). « Early identification of potentially salvageable tissue with MRI-based predictive algorithms after experimental ischemic stroke ». Dans : *Journal of Cerebral Blood Flow and Metabolism* 33.1. URL : <http://dx.doi.org/10.1038/jcbfm.2013.51>.
- BOYD, K., K. ENG et C. PAGE (2013). « Area under the Precision-Recall Curve : Point Estimates and Confidence Intervals ». Dans : *Machine Learning and Knowledge Discovery in Databases*. Lecture Notes in Computer Science 8190, p. 451–466. URL : http://dx.doi.org/10.1007/978-3-642-40994-3_29.
- BRAUN, J., J. BERNARDING, H.-C. KOENNECKE, K.-J. WOLF et T. TOLXDORFF (2002). « Feature-based, Automated Segmentation of Cerebral Infarct Patterns Using T2- and Diffusion-weighted Imaging ». Dans :

- Computer Methods in Biomechanics and Biomedical Engineering* 5.6, p. 411–420. URL : <http://dx.doi.org/10.1080/1025584021000011082>.
- BROTT, T, H. P. ADAMS, C. P. OLINGER, J. R. MARLER, W. G. BARSAN, J. BILLER, J. SPILKER, R. HOLLERAN, R. EBERLE et V. HERTZBERG (1989). « Measurements of acute cerebral infarction : a clinical examination scale. » Dans : *Stroke* 20.7, p. 864–70. URL : <http://dx.doi.org/10.1161/01.STR.20.7.864>.
- CAI, D., X. HE et J. HAN (2007). « Spectral Regression for Efficient Regularized Subspace Learning ». Dans : *IEEE 11th International Conference on Computer Vision*, p. 1–8. URL : <http://dx.doi.org/10.1109/iccvc.2007.4408855>.
- CAMPBELL, B. C. V., A. PURUSHOTHAM, S. CHRISTENSEN, P. M. DESMOND, Y. NAGAKANE, M. W. PARSONS, M. G. LANSBERG, M. MLYNASH, M. STRAKA, D. A. DE SILVA, J.-M. OLIVOT, R. BAMMER, G. W. ALBERS, G. A. DONNAN et S. M. DAVIS (2012). « The infarct core is well represented by the acute diffusion lesion : sustained reversal is infrequent ». Dans : *Journal of Cerebral Blood Flow and Metabolism* 32.1, p. 50–56. URL : <http://dx.doi.org/10.1038/jcbfm.2011.102>.
- CARANO, R. A., F. LI, K. IRIE, K. G. HELMER, M. D. SILVA, M. FISHER et C. H. SOTAK (2000). « Multispectral analysis of the temporal evolution of cerebral ischemia in the rat brain ». Dans : *Journal of Magnetic Resonance Imaging* 12.6, p. 842–858. URL : [http://dx.doi.org/10.1002/1522-2586\(200012\)12:6<842::AID-JMRI7>3.0.CO;2-5](http://dx.doi.org/10.1002/1522-2586(200012)12:6<842::AID-JMRI7>3.0.CO;2-5).
- CARRERA, E., P. S. JONES, J. A. ALAWNEH, I. KLÉRKE MIKKELSEN, T.-H. CHO, S. SIEMONSEN, J. V. GUADAGNO, K. MOURIDSEN, L. RIBE, N. HJORT, T. D. FRYER, T. A. CARPENTER, F. I. AIGBIRHIO, J. FIEHLER, N. NIGHOGHOSSIAN, E. A. WARBURTON, L. ØSTERGAARD et J.-C. BARON (2011). « Predicting Infarction Within the Diffusion-Weighted Imaging Lesion : Does the Mean Transit Time Have Added Value ? » Dans : *Stroke* 42.6, p. 1602–1607. URL : <http://dx.doi.org/10.1161/STROKEAHA.110.606970>.
- CASALS, J. B., N. C. PIERI, M. L. FEITOSA, A. C. ERCOLIN, K. C. ROBALLO, R. S. BARRETO, F. F. BRESSAN, D. S. MARTINS, M. A. MIGLINO et C. E. AMBRÓSIO (2011). « The Use of Animal Models for Stroke Research : A Review ». Dans : *Comparative Medicine* 61.4, p. 305–313. URL : <http://www.ingentaconnect.com/content/aalas/cm/2011/00000061/00000004/art00002>.
- CELEUX, G., F. FORBES et N. PEYRARD (2003). « EM procedures using mean field-like approximations for Markov model-based image segmentation ». Dans : *Pattern Recognition* 36.1, p. 131–144. URL : [http://dx.doi.org/10.1016/S0031-3203\(02\)00027-4](http://dx.doi.org/10.1016/S0031-3203(02)00027-4).
- CELEUX, G. et G. GOVAERT (1991). « A classification EM algorithm for clustering and two stochastic versions ». Dans : *Computational Statistics and Data Analysis* 14.3, p. 315–332. URL : [http://dx.doi.org/10.1016/0167-9473\(92\)90042-e](http://dx.doi.org/10.1016/0167-9473(92)90042-e).
- CHALELA, J. A., C. S. KIDWELL, L. M. NENTWICH, M. LUBY, J. A. BUTMAN, A. M. DEMCHUK, M. D. HILL, N. PATRONAS, L. LATOUR et S. WARACH (2007). « Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke : a prospective comparison ». Dans : *The Lancet* 369.9558, p. 293–298. URL : [http://dx.doi.org/10.1016/s0513-5117\(08\)79190-7](http://dx.doi.org/10.1016/s0513-5117(08)79190-7).
- CHAWLA, M., S. SHARMA, J. SIVASWAMY et L. KISHORE (2009). « A method for automatic detection and classification of stroke from brain CT images ». Dans : p. 3581–3584. URL : <http://dx.doi.org/10.1109/IEMBS.2009.5335289>.

- CHRISTENSEN, S., K. MOURIDSEN, O. WU, N. HJORT, H. KARSTOFT, G. THOMALLA, J. RÖTHER, J. FIEHLER, T. KUCINSKI et L. ØSTERGAARD (2009). « Comparison of 10 Perfusion MRI Parameters in 97 Sub-6-Hour Stroke Patients Using Voxel-Based Receiver Operating Characteristics Analysis ». Dans : *Stroke* 40.6, p. 2055–2061. URL : <http://dx.doi.org/10.1161/STROKEAHA.108.546069>.
- CLARK, W. M., G. W. ALBERS, K. P. MADDEN, S. HAMILTON et for the Thrombolytic Therapy in ACUTE ISCHEMIC STROKE STUDY INVESTIGATORS (2000). « The rtPA (Alteplase) 0- to 6-Hour Acute Stroke Trial, Part A : Results of a Double-Blind, Placebo-Controlled, Multicenter Study ». Dans : *Stroke* 31.4, p. 811–816. URL : <http://dx.doi.org/10.1161/01.STR.31.4.811>.
- CLARK, W., S. WISSMAN, G. ALBERS, J. JHAMANDAS, K. MADDEN et S. HAMILTON (1999). « Recombinant tissue-type plasminogen activator (alteplase) for ischemic stroke 3 to 5 hours after symptom onset : The atlantis study : a randomized controlled trial ». Dans : *Journal of the American Medical Association* 282.21, p. 2019–2026. URL : <http://dx.doi.org/10.1001/jama.282.21.2019>.
- CLAS, P., S. GROESCHEL et M. WILKE (2012). « A Semi-Automatic Algorithm for Determining the Demyelination Load in Metachromatic Leukodystrophy ». Dans : *Academic Radiology* 19.1, p. 26 –34. URL : <http://dx.doi.org/10.1016/j.acra.2011.09.008>.
- COMANICIU, D. et P. MEER (2002). « Mean Shift : A Robust Approach Toward Feature Space Analysis ». Dans : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.5, p. 603–619. URL : <http://dx.doi.org/10.1109/34.1000236>.
- CONTIN, L., C. BEER, M. BYNEVELT, H. WITTSACK et G. GARRIDO (2010). « Semi-automatic segmentation of core and penumbra regions in acute ischemic stroke : preliminary results. » Dans : *17th International Conference on Systems, Signals and Image Processing*. URL : <http://dx.doi.org/10.1161/01.STR.31.4.811>.
- DALCA, A. V., R. SRIDHARAN, L. CLOONAN, K. M. FITZPATRICK, A. KANAKIS, K. L. FURIE, J. ROSAND, O. WU, M. SABUNCU, N. S. ROST et P. GOLLAND (2014). « Segmentation of Cerebrovascular Pathologies in Stroke Patients with Spatial and Shape Priors ». Dans : *Med Image Comput Comput Assist Interv.* 17.2, 773–780. URL : http://dx.doi.org/10.1007/978-3-319-10470-6_96.
- DASTIDAR, P., T. HEINONEN, J.-P. AHONEN, M. JEHKONEN et G. MOLNÁR (2000). « Volumetric measurements of right cerebral hemisphere infarction : use of a semiautomatic MRI segmentation technique ». Dans : *Computers in Biology and Medicine* 30.1, p. 41 –54. URL : [http://dx.doi.org/10.1016/S0010-4825\(99\)00022-0](http://dx.doi.org/10.1016/S0010-4825(99)00022-0).
- DAVIS, J. et M. GOADRICH (2006). « The Relationship Between Precision-Recall and ROC Curves ». Dans : *Proceedings of the 23rd International Conference on Machine Learning*. ICML '06, p. 233–240. URL : <http://dx.doi.org/10.1145/1143844.1143874>.
- DEMPSTER, A. P., N. M. LAIRD et D. B. RUBIN (1977). « Maximum likelihood from incomplete data via the EM algorithm ». Dans : *Journal of the royal statistical society. Series B (Methodological)* 39.1, p. 1–38. URL : <http://web.mit.edu/6.435/www/Dempster77.pdf>.
- DESCOMBES, X., R. MORRIS et J. ZERUBIA (1996). *Quelques améliorations à la segmentation d'images bayésiennes*. Rap. tech. URL : <https://hal.inria.fr/inria-00073781/file/RR-2916.pdf>.

- DESPOTOVIĆ, I., B. GOOSSENS et W. PHILIPS (2015). « MRI Segmentation of the Human Brain : Challenges, Methods, and Applications ». Dans : *Computational and Mathematical Methods in Medicine*. URL : <http://dx.doi.org/10.1155/2015/450341>.
- DWYER, M. G., N. BERGLAND, E. SALUSTE, J. SHARMA, Z. JAISANI, J. DURFEE, N. ABDELRAHMAN, A. MINAGAR, R. HOQUE, F. E. MUNSCHAUER III et R. ZIVADINOV (2008). « Application of hidden Markov random field approach for quantification of perfusion/diffusion mismatch in acute ischemic stroke ». Dans : *Neurological Research* 30.8, p. 827–834. URL : <http://dx.doi.org/10.1179/174313208X340987>.
- EDDELBUEITTEL, D. et R. FRANCOIS (2011). « Rcpp : Seamless R and C++ Integration ». Dans : *Journal of Statistical Software* 40.8, p. 1–18. URL : <http://www.jstatsoft.org/v40/i08>.
- FALCAO, A. L. E., D. C. REUTENS, R. MARKUS, M. KOGA, S. J. READ, H. TOCHON-DANGUY, J. SACHINIDIS, D. W. HOWELLS et G. A. DONNAN (2004). « The resistance to ischemia of white and gray matter after stroke ». Dans : *Annals of Neurology* 56.5, p. 695–701. URL : <http://dx.doi.org/10.1002/ana.20265>.
- FAWCETT, T. (2006). « An Introduction to ROC Analysis ». Dans : *Pattern Recognition Letters* 27.8, p. 861–874. URL : <http://dx.doi.org/10.1016/j.patrec.2005.10.010>.
- FENG, D., L. TIERNEY et V. MAGNOTTA (2012). « MRI Tissue Classification Using High-Resolution Bayesian Hidden Markov Normal Mixture Models ». Dans : *Journal of the American Statistical Association* 107.497, p. 102–119. URL : <http://dx.doi.org/10.1198/jasa.2011.ap09529>.
- FENG, D. et L. TIERNEY (2011). « mritc : A Package for MRI Tissue Classification ». Dans : *Journal of Statistical Software* 44. URL : www.jstatsoft.org/article/view/v044i07/v44i07.pdf.
- FIEHLER, J., M. FOTH, T. KUCINSKI, R. KNAB, M. von BEZOLD, C. WEILLER, H. ZEUMER et J. RÖTHER (2002). « Severe ADC Decreases Do Not Predict Irreversible Tissue Damage In Humans ». Dans : *Stroke* 33.1, p. 79–86. URL : <http://dx.doi.org/10.1161/hs0102.100884>.
- FORBES, F. et G. FORT (2007). « Combining Monte Carlo and Mean-Field-Like Methods for Inference in Hidden Markov Random Fields ». Dans : *IEEE Transactions on Image Processing* 16.3, p. 824–837. URL : <http://dx.doi.org/10.1109/TIP.2006.891045>.
- FORBES, F., S. DOYLE, D. GARCIA-LORENZO, C. BARILLOT et M. DOJAT (2010a). « A weighted Multi-sequence Markov model for brain lesion segmentation ». Dans : *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, p. 225–232. URL : <http://mistis.inrialpes.fr/~forbes/PAPERS/aistats2010-FINAL.pdf>.
- (2010b). « Adaptive weighted fusion of multiple MR sequences for brain lesion segmentation ». Dans : *IEEE International Symposium on Biomedical Imaging (ISBI)*, p. 69–72. URL : <http://dx.doi.org/10.1109/ISBI.2010.5490413>.
- FORKERT, N., P. KAESEMAN, A. TRESZL, S. SIEMONSEN, B. CHENG, H. HANDELS, J. FIEHLER et G. THOMALLA (2013). « Comparison of 10 TTP and Tmax Estimation Techniques for MR Perfusion-Diffusion Mismatch Quantification in Acute Stroke ». Dans : *American Journal of Neuroradiology* 34.9, p. 1697–1703. URL : <http://dx.doi.org/10.3174/ajnr.A3460>.

- FRIEDMAN, J., T. HASTIE et R. TIBSHIRANI (2000). « Additive logistic regression : a statistical view of boosting ». Dans : *The Annals of Statistics* 28.2, p. 337–407. URL : <http://dx.doi.org/10.1214/aos/1016218223>.
- FRINDEL, C., M. C. ROBINI et D. ROUSSEAU (2015a). « A 3D spatio-temporal deconvolution approach for MR perfusion in the brain ». Dans : *Medical Image Analysis* 18, p. 144–160. URL : <http://dx.doi.org/10.1016/j.media.2013.10.004>.
- FRINDEL, C., A. ROUANET, M. GIACALONE, T.-H. CHO, L. ØSTERGAARD, J. FIEHLER, S. PEDRAZA, J.-C. BARON, M. WIART, Y. BERTHEZÈNE, N. NIGHOGHOSSIAN et D. ROUSSEAU (2015b). « Validity of Shape as a Predictive Biomarker of Final Infarct Volume in Acute Ischemic Stroke ». Dans : *Stroke*. URL : <http://dx.doi.org/10.1161/STROKEAHA.114.008046>.
- GAETAN, C. et X. GUYON (2010). *Spatial Statistics and Modeling*. Springer. URL : <http://www.springer.com/us/book/9780387922560>.
- GEMAN, S. et D. GEMAN (1984). « Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images ». Dans : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6.6, p. 721–741. URL : <http://dx.doi.org/10.1109/TPAMI.1984.4767596>.
- GENOLINI, C. (2009). *Petit Manuel de S4 : Programmation Orientée Objet sous R*. URL : <http://cran.r-project.org/doc/contrib/Genolini-PetitManuelDeS4.pdf>.
- GHOSH, N., Y. SUN, B. BHANU, S. ASHWAL et A. OBENAU (2014). « Automated detection of brain abnormalities in neonatal hypoxia ischemic injury from MR images ». Dans : *Medical Image Analysis* 18.7, p. 1059–1069. URL : <http://dx.doi.org/10.1016/j.media.2014.05.002>.
- GILLEBERT, C. R., G. W. HUMPHREYS et D. MANTINI (2014). « Automated delineation of stroke lesions using brain CT images ». Dans : *NeuroImage : Clinical* 4, p. 540–548. URL : <http://dx.doi.org/10.1016/j.nicl.2014.03.009>.
- GOADRICH, M., L. OLIPHANT et J. SHAVLIK (2004). « Learning Ensembles of First-Order Clauses for Recall-Precision Curves : A Case Study ». Dans : *Biomedical Information Extraction*, p. 98–115. URL : http://dx.doi.org/10.1007/978-3-540-30109-7_11.
- GOTTARDO, R., J. BESAG, M. STEPHENS et A. MURUA (2006). « Probabilistic segmentation and intensity estimation for microarray images ». Dans : *Biostatistics* 7.1, p. 85–99. URL : <http://dx.doi.org/10.1093/biostatistics/kxi042>.
- GOTTRUP, C., K. THOMSEN, P. LOCHT, O. WU, A. G. SORESENSEN, W. J. KOROSHETZ et L. ØSTERGAARD (2005). « Applying instance-based techniques to prediction of final outcome in acute stroke ». Dans : *Artificial Intelligence in Medicine* 33.3, p. 223–236. URL : <http://dx.doi.org/10.1016/j.artmed.2004.06.003>.
- GRANDIN, C. B., T. P. DUPREZ, A. M. SMITH, C. OPPENHEIM, A. PEETERS, A. R. ROBERT et G. COSNARD (2002). « Which MR-derived Perfusion Parameters are the Best Predictors of Infarct Growth in Hyperacute Stroke? Comparative Study between Relative and Quantitative Measurements ». Dans : *Radiology* 223.2, p. 361–370. URL : <http://dx.doi.org/10.1148/radiol.2232010673>.

- GRIMES, D. A. et K. F. SCHULZ (2002). « Uses and abuses of screening tests ». Dans : *The Lancet* 359.9309, p. 881–884. URL : [http://dx.doi.org/10.1016/S0140-6736\(02\)07948-5](http://dx.doi.org/10.1016/S0140-6736(02)07948-5).
- GUNAWARDANA, A. et W. BYRNE (2005). « Convergence Theorems for Generalized Alternating Minimization Procedures ». Dans : *Journal of Machine Learning Research* 6, p. 2049–2073. URL : <http://dl.acm.org/citation.cfm?id=1046920.1194913>.
- HAAN, B. de, P. CLAS, H. JUENGER, M. WILKE et H.-O. KARNATH (2015). « Fast semi-automated lesion demarcation in stroke ». Dans : *NeuroImage : Clinical* 9, p. 69–74. URL : <http://dx.doi.org/10.1016/j.nicl.2015.06.013>.
- HACKE, W., M. KASTE, E. BLUHMKI, M. BROZMAN, A. DÁVALOS, D. GUIDETTI, V. LARRUE, K. R. LEES, Z. MEDEGHRI, T. MACHNIG, D. SCHNEIDER, R. von KUMMER, N. WAHLGREN et D. TONI (2008). « Thrombolysis with Alteplase 3 to 4.5 Hours after Acute Ischemic Stroke ». Dans : *New England Journal of Medicine* 359.13, p. 1317–1329. URL : <http://dx.doi.org/10.1056/NEJMoa0804656>.
- HACKE, W., A. J. FURLAN, Y. AL-RAWI, A. DAVALOS, J. B. FIEBACH, F. GRUBER, M. KASTE, L. J. LIPKA, S. PEDRAZA, P. A. RINGLEB, H. A. ROWLEY, D. SCHNEIDER, L. H. SCHWAMM, J. S. LEAL, M. SÖHNGEN, P. A. TEAL, K. WILHELM-OGUNBIYI, M. WINTERMARK et S. WARACH (2009). « Intravenous desmoteplase in patients with acute ischaemic stroke selected by MRI perfusion–diffusion weighted imaging or perfusion CT (DIAS-2) : a prospective, randomised, double-blind, placebo-controlled study ». Dans : *The Lancet Neurology* 8.2, p. 141–150. URL : [http://dx.doi.org/10.1016/S1474-4422\(08\)70267-9](http://dx.doi.org/10.1016/S1474-4422(08)70267-9).
- HANSEN, M. B., K. NAGENTHIRAJA, L. R. RIBE, K. H. DUPONT, L. ØSTERGAARD et K. MOURIDSEN (2015). « Automated estimation of salvageable tissue : Comparison with expert readers ». Dans : *Journal of Magnetic Resonance Imaging*. URL : <http://dx.doi.org/10.1002/jmri.24963>.
- HAUTE AUTORITÉ DE SANTÉ (2009). « Accident vasculaire cérébral : prise en charge précoce ». Dans : URL : http://www.has-sante.fr/portail/upload/docs/application/pdf/2009-07/avc_prise_en_charge_precoce_-_argumentaire.pdf.
- HE, H. et E. GARCIA (2009). « Learning from Imbalanced Data ». Dans : *IEEE Transactions on Knowledge and Data Engineering* 21.9, p. 1263–1284. URL : <http://dx.doi.org/10.1109/TKDE.2008.239>.
- HEVIA-MONTIEL, N., J. JIMENEZ-ALANIZ, V. MEDINA-BANUELOS, O. YANEZ-SUAREZ, C. ROSSO, Y. SAMSON et S. BAILLET (2007). « Robust Nonparametric Segmentation of Infarct Lesion from Diffusion-Weighted MR Images ». Dans : *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, p. 2102–2105. URL : <http://dx.doi.org/10.1109/IEMBS.2007.4352736>.
- HÂNCU, A., I. RĂȘANU et G. BUTOI (2012). *Advances in Brain Imaging. Chapter 12 : White Matter Changes in Cerebrovascular Disease : Leukoaraiosis*. InTech. URL : <http://dx.doi.org/10.5772/30565>.
- HOA, D., P. TAOUREL, A. MICHEAU, G. GAHIDE et E. LE BARS (2008). *L’IRM pas à pas*. Spécialités Médicales. Sauramps Médical. URL : <https://www.imaios.com/fr/e-Cours/e-MRI>.
- HUANG, S., Q. SHEN et T. Q. DUONG (2010). « Artificial neural network prediction of ischemic tissue fate in acute stroke imaging ». Dans : *Journal of Cerebral Blood Flow and Metabolism* 30.9, p. 1661–1670. URL : <http://dx.doi.org/10.1038/jcbfm.2010.56>.

- HUANG, S., Q. SHEN et T. Q. DUONG (2011). « Quantitative prediction of acute ischemic tissue fate using support vector machine ». Dans : *Brain Research* 1405, p. 77–84. URL : <http://dx.doi.org/10.1016/j.brainres.2011.05.066>.
- HUGHES, J., M. HARAN et P. C. CARAGEA (2011). « Autologistic models for binary data on a lattice ». Dans : *Environmetrics* 22.7, p. 857–871. URL : <http://dx.doi.org/10.1002/env.1102>.
- IST-3 COLLABORATIVE GROUP (2012). « The benefits and harms of intravenous thrombolysis with recombinant tissue plasminogen activator within 6 h of acute ischaemic stroke ». Dans : *The Lancet* 379.9834, p. 2352–2363. URL : [http://dx.doi.org/10.1016/S0140-6736\(12\)60768-5](http://dx.doi.org/10.1016/S0140-6736(12)60768-5).
- JACOBS, M. A., R. A. KNIGHT, H. SOLTANIAN-ZADEH, Z. G. ZHENG, A. V. GOUSSEV, D. J. PECK, J. P. WINDHAM et M. CHOPP (2000). « Unsupervised segmentation of multiparameter MRI in experimental cerebral ischemia with comparison to T2, diffusion, and ADC MRI parameters and histopathological validation ». Dans : *Journal of Magnetic Resonance Imaging* 11.4, p. 425–437. URL : [http://dx.doi.org/10.1002/\(SICI\)1522-2586\(200004\)11:4<425::AID-JMRI11>3.0.CO;2-0](http://dx.doi.org/10.1002/(SICI)1522-2586(200004)11:4<425::AID-JMRI11>3.0.CO;2-0).
- JACOBS, M. A., Z. G. ZHANG, R. A. KNIGHT, H. SOLTANIAN-ZADEH, A. V. GOUSSEV, D. J. PECK et M. CHOPP (2001a). « A Model for Multiparametric MRI Tissue Characterization in Experimental Cerebral Ischemia With Histological Validation in Rat : Part 1 ». Dans : *Stroke* 32.4, p. 943–949. URL : <http://dx.doi.org/10.1161/01.STR.32.4.943>.
- JACOBS, M. A., P. MITSIAS, H. SOLTANIAN-ZADEH, S. SANTHAKUMAR, A. GHANEI, R. HAMMOND, D. J. PECK, M. CHOPP et S. PATEL (2001b). « Multiparametric MRI Tissue Characterization in Clinical Stroke With Correlation to Clinical Outcome : Part 2 ». Dans : *Stroke* 32.4, p. 950–957. URL : <http://dx.doi.org/10.1161/01.STR.32.4.950>.
- JAHNH, G.-H., K.-L. LI, L. ØSTERGAARD et F. CALAMANTE (2014). « Perfusion Magnetic Resonance Imaging : A Comprehensive Update on Principles and Techniques ». Dans : *Korean Journal Radiology* 5.15, p. 554–577. URL : <http://dx.doi.org/10.3348/kjr.2014.15.5.554>.
- JAMES, J. R., K. K. YODER, O. OSUNTOKUN, A. KALNIN, A. BRUNO et E. D. MORRIS (2006). « A supervised method for calculating perfusion/diffusion mismatch volume in acute ischemic stroke ». Dans : *Computers in Biology and Medicine* 36.11, p. 1268–1287. URL : <http://dx.doi.org/10.1016/j.combiomed.2005.05.007>.
- JAUCH, E. C., J. L. SAVER, H. P. ADAMS, A. BRUNO, J. B. CONNORS, B. M. DEMAERSCHALK, P. KHATRI, P. W. McMULLAN, A. I. QURESHI, K. ROSENFELD, P. A. SCOTT, D. R. SUMMERS, D. Z. WANG, M. WINTERMARK et H. YONAS (2013). « Guidelines for the Early Management of Patients With Acute Ischemic Stroke : A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association ». Dans : *Stroke* 44.3, p. 870–947. URL : <http://dx.doi.org/10.1161/STR.0b013e318284056a>.
- JONSDOTTIR, K. Y., L. ØSTERGAARD et K. MOURIDSEN (2009). « Predicting Tissue Outcome From Acute Stroke Magnetic Resonance Imaging : Improving Model Performance by Optimal Sampling of Training Data ». Dans : *Stroke* 40.9, p. 3006–3011. URL : <http://dx.doi.org/10.1161/STROKEAHA.109.552216>.
- KABIR, Y., M. DOJAT, B. SCHERRER, F. FORBES et C. GARBAY (2007). « Multimodal MRI segmentation of ischemic stroke lesions ». Dans : *29th Annual International Conference of the IEEE* 1, p. 1595–1598. URL : <http://dx.doi.org/10.1109/IEMBS.2007.4352610>.

- KANEKAR, S. G., T. ZACHARIA et R. ROLLER (2012). « Imaging of Stroke : Part 2, Pathophysiology at the Molecular and Cellular Levels and Corresponding Imaging Changes ». Dans : *American Journal of Roentgenology* 198, p. 63–74. URL : <http://dx.doi.org/10.2214/ajr.10.7312>.
- KIDWELL, C. S., J. R. ALGER et J. L. SAVER (2003). « Beyond Mismatch : Evolving Paradigms in Imaging the Ischemic Penumbra With Multimodal Magnetic Resonance Imaging ». Dans : *Stroke* 34.1, p. 2729–2735. URL : <http://dx.doi.org/10.1161/01.str.0000097608.38779.cc>.
- KIDWELL, C. S., M. WINTERMARK, D. A. DE SILVA, T. J. SCHAEWE, R. JAHAN, S. STARKMAN, T. JOVIN, J. HOM, M. JUMAA, J. SCHREIER, J. GORNBEIN, D. S. LIEBESKIND, J. R. ALGER et J. L. SAVER (2013). « Multiparametric MRI and CT Models of Infarct Core and Favorable Penumbra Imaging Patterns in Acute Ischemic Stroke ». Dans : *Stroke* 44.1, p. 73–79. URL : <http://dx.doi.org/10.1161/STROKEAHA.112.670034>.
- KISSELA, B. M., J. C. KHOURY, K. ALWELL, C. J. MOOMAW, D. WOO, O. ADEOYE, M. L. FLAHERTY, P. KHATRI, S. FERIOLI, F. DE LOS RIOS LA ROSA, J. P. BRODERICK et D. O. KLEINDORFER (2012). « Age at stroke : temporal trends in stroke incidence in a large, biracial population. » Dans : *Neurology* 79.17, p. 1781–7. URL : <http://dx.doi.org/10.1212/wnl.0b013e318270401d>.
- KLËPPEL, S., A. ABDULKADIR, S. HADJIDEMETRIOU, S. ISSLEIB, L. FRINGS et al. (2011). « A comparison of different automated methods for the detection of white matter lesions in MRI data ». Dans : *NeuroImage* 57.2, p. 416–422. URL : <http://dx.doi.org/10.1016/j.neuroimage.2011.04.053>.
- KLIER, T. H. et D. MCMILLEN (2005). « Clustering of auto supplier plants in the U.S. : GMM spatial logit for large samples ». Dans : FRB Chicago Working Paper. URL : <http://dx.doi.org/10.2139/ssrn.870611>.
- KNAUS, J. (2013). *snowfall : Easier cluster computing (based on snow)*. R package version 1.84-6. URL : <http://CRAN.R-project.org/package=snowfall>.
- KOGA, M., D. C. REUTENS, P. WRIGHT, T. PHAN, R. MARKUS, B. PEDREIRA, G. FITT, I. LIM et G. A. DONNAN (2005). « The Existence and Evolution of Diffusion–Perfusion Mismatched Tissue in White and Gray Matter After Acute Stroke ». Dans : *Stroke* 36.10, p. 2132–2137. URL : <http://dx.doi.org/10.1161/01.STR.0000181066.23213.8f>.
- KOHLI, P., L. LADICKY et P. H. TORR (2009). « Robust Higher Order Potentials for Enforcing Label Consistency ». Dans : *International Journal of Computer Vision* 82.3, p. 302–324. URL : <http://dx.doi.org/10.1007/s11263-008-0202-0>.
- LANSBERG, M. G., M. STRAKA, S. KEMP, M. MLYNASH, L. R. WECHSLER, T. G. JOVIN, M. J. WILDER, H. L. LUTSEP, T. J. CZARTOSKI, R. A. BERNSTEIN, C. W. CHANG, S. WARACH, F. FAZEKAS, M. INOUE, A. TIPIRNENI, S. A. HAMILTON, G. ZAHARCHUK, M. P. MARKS, R. BAMMER et G. W. ALBERS (2012). « MRI profile and response to endovascular reperfusion after stroke (DEFUSE 2) : a prospective cohort study ». Dans : *The Lancet Neurology* 11.10, p. 860–867. URL : [http://dx.doi.org/10.1016/S1474-4422\(12\)70203-X](http://dx.doi.org/10.1016/S1474-4422(12)70203-X).
- LARRUE, V., R. von KUMMER, G. del ZOPPO et E. BLUHMKI (1997). « Hemorrhagic Transformation in Acute Ischemic Stroke : Potential Contributing Factors in the European Cooperative Acute Stroke Study ». Dans : *Stroke* 28.1, p. 957–960. URL : <http://dx.doi.org/10.1161/01.STR.28.5.957>.

- LEES, K. R., E. BLUHMKI, R. von KUMMER, T. G. BROTT, D. TONI, J. C. GROTTA, G. W. ALBERS, M. KASTE, J. R. MARLER, S. A. HAMILTON, B. C. TILLEY, S. M. DAVIS, G. A. DONNAN et W. HACKE (2010). « Time to treatment with intravenous alteplase and outcome in stroke : an updated pooled analysis of ECASS, ATLANTIS, NINDS, and EPITHET trials ». Dans : *The Lancet* 375.9727, p. 1695 –1703. URL : [http://dx.doi.org/10.1016/S0140-6736\(10\)60491-6](http://dx.doi.org/10.1016/S0140-6736(10)60491-6).
- LEVY-COOPERMAN, N., J. RAMIREZ, N. J. LOBAUGH et S. E. BLACK (2008). « Misclassified Tissue Volumes in Alzheimer Disease Patients With White Matter Hyperintensities : Importance of Lesion Segmentation Procedures for Volumetric Analysis ». Dans : *Stroke* 39.4, p. 1134–1141. URL : <http://dx.doi.org/10.1161/strokeaha.107.498196>.
- LI, M., L. AI, H. HE, Z. ZHENG, B. LV, W. LI, J. YI et X. CHEN (2009). « Segmentation of infarct in acute ischemic stroke from MR apparent diffusion coefficient and trace-weighted images ». Dans : *Proceedings of SPIE* 7497. URL : <http://dx.doi.org/10.1117/12.832535>.
- LI, S. Z. (2009). *Markov Random Field Modeling in Image Analysis*. Springer. URL : <http://www.springer.com/us/book/9781848002784>.
- LI, W., J. TIAN, E. LI et J. DAI (2004). « Robust unsupervised segmentation of infarct lesion from diffusion tensor MR images using multiscale statistical classification and partial volume voxel reclassification ». Dans : *NeuroImage* 23.4, p. 1507 –1518. URL : <http://dx.doi.org/10.1016/j.neuroimage.2004.08.009>.
- LLADÓ, X., A. OLIVER, M. CABEZAS, J. FREIXENET, J. C. VILANOVA, A. QUILES, L. VALLS, L. RAMIÓ-TORRENTÍ et I. ROVIRA (2012). « Segmentation of Multiple Sclerosis Lesions in Brain MRI : A Review of Automated Approaches ». Dans : *Information Sciences* 186.1, p. 164–185. URL : <http://dx.doi.org/10.1016/j.ins.2011.10.011>.
- LOH, W.-Y. (2011). « Classification and regression trees ». Dans : *Wiley Interdisciplinary Reviews : Data Mining and Knowledge Discovery* 1.1, p. 14–23. URL : <http://dx.doi.org/10.1002/widm.8>.
- LU, M., P. D. MITSIAS, J. R. EWING, H. SOLTANIAN-ZADEH, H. BAGHER-EBADIAN, Q. ZHAO, N. OJA-TEBBE, S. C. PATEL et M. CHOPP (2005). « Predicting final infarct size using acute and subacute multiparametric MRI measurements in patients with ischemic stroke ». Dans : *Journal of Magnetic Resonance Imaging* 21.5, p. 495–502. URL : <http://dx.doi.org/10.1002/jmri.20313>.
- MAH, Y.-H., R. JAGER, C. KENNARD, M. HUSAIN et P. NACHEV (2014). « A new method for automated high-dimensional lesion segmentation evaluated in vascular injury and applied to the human occipital lobe ». Dans : *Cortex* 56, p. 51 –63. URL : <http://dx.doi.org/10.1016/j.cortex.2012.12.008>.
- MAIER, O., M. WILMS, J. von der GABLENTZ, U. M. KRÄMER, T. F. MÜNTE et H. HANDELS (2015). « Extra Tree forests for sub-acute ischemic stroke lesion segmentation in MR sequences ». Dans : *Journal of Neuroscience Methods* 240, p. 89 –100. URL : <http://dx.doi.org/10.1016/j.jneumeth.2014.11.011>.
- MALDJIAN, J. A., J. CHALELA, S. E. KASNER, D. LIEBESKIND et J. A. DETRE (2001). « Automated CT Segmentation and Analysis for Acute Middle Cerebral Artery Stroke ». Dans : *American Journal of Neuroradiology* 22.6, p. 1050–1055. URL : <http://www.ajnr.org/content/22/6/1050.abstract>.
- MANNING, C. D., P. RAGHAVAN et H. SCHÜTZE (2008). *Introduction to Information Retrieval (Chapter 8)*. Cambridge University Press. URL : <http://nlp.stanford.edu/IR-book/pdf/08eval.pdf>.

- MARTEL, A., S. ALLDER, G. DELAY, P. MORGAN et A. MOODY (1999). « Measurement of Infarct Volume in Stroke Patients Using Adaptive Segmentation of Diffusion Weighted MR Images ». Dans : *Medical Image Computing and Computer-Assisted Intervention – MICCAI’99* 1679, p. 22–31. URL : http://dx.doi.org/10.1007/10704282_3.
- MATESIN, M., S. LONCARIC et D. PETRAVIC (2001). « A rule-based approach to stroke lesion analysis from CT brain images ». Dans : *Image and Signal Processing and Analysis, 2001. ISPA 2001. Proceedings of the 2nd International Symposium on*, p. 219–223. URL : <http://dx.doi.org/10.1109/ISPA.2001.938631>.
- MEILŪNAS, M., A. UŠINSKAS, R. KIRVAITIS et R. DOBROVOLSIS (2003). « Automatic contouring of segmented human brain ischemic stroke region on CT images ». Dans : *Mathematical Modelling and Analysis* 8.1, p. 43–50. URL : <http://dx.doi.org/10.1080/13926292.2003.9637209>.
- MITRA, J., P. BOURGEAT, J. FRIPP, S. GHOSE, S. ROSE, O. SALVADO, A. CONNELLY, B. CAMPBELL, S. PALMER, G. SHARMA, S. CHRISTENSEN et L. CAREY (2014). « Lesion segmentation from multimodal MRI using random forest following ischemic stroke ». Dans : *NeuroImage* 98.0, p. 324–335. URL : <http://dx.doi.org/10.1016/j.neuroimage.2014.04.056>.
- MOLINARO, A. M., R. SIMON et R. M. PFEIFFER (2005). « Prediction error estimation : a comparison of resampling methods ». Dans : *Bioinformatics* 21.15, p. 3301–3307. URL : <http://dx.doi.org/10.1093/bioinformatics/bti499>.
- MORRIS, R., X. DESCOMBES et J. ZERUBIA (1996). « The Ising/Potts model is not well suited to segmentation tasks ». Dans : *Digital Signal Processing Workshop Proceedings*, p. 263–266. URL : <http://dx.doi.org/10.1109/DSPWS.1996.555511>.
- MOSKOWITZ, M. A., E. H. LO et C. IADECOLA (2010). « The Science of Stroke : Mechanisms in Search of Treatments ». Dans : *Neuron* 67.2, p. 181–198. URL : <http://dx.doi.org/10.1016/j.neuron.2010.07.002>.
- MOURIDSEN, K., O. WU, W. J. KOROSHETZ, A. G. SORENSEN et L. ØSTERGAARD (2004). « Optimal Parameter Choice in Predicting Final Outcome in Acute Stroke ». Dans : *Proceedings of the International Society for Magnetic Resonance in Medicine* 11. URL : <http://cds.ismrm.org/ismrm-2004/Files/000410.pdf>.
- MOURIDSEN, K., K. NAGENTHIRAJA, K. Y. JONSDOTTIR, L. R. RIBE, A. B. NEUMANN, N. HJORT et L. ØSTERGAARD (2013). « Acute Stroke : Automatic Perfusion Lesion Outlining Using Level Sets ». Dans : *Radiology* 269.2, p. 404–412. URL : <http://dx.doi.org/10.1148/radiology.13121622>.
- MOURIDSEN, K., M. B. HANSEN, L. ØSTERGAARD et S. N. JESPERSEN (2014). « Reliable estimation of capillary transit time distributions using DSC-MRI ». Dans : *Journal of Cerebral Blood Flow and Metabolism* 34.9, p. 1511–1521. URL : <http://dx.doi.org/10.1038/jcbfm.2014.111>.
- MUSCHELLI, J. (2015). *fslr : Wrapper Functions for FSL (FMRIB Software Library) from Functional MRI of the Brain (FMRIB)*. R package version 1.4.4. URL : <http://CRAN.R-project.org/package=fslr>.
- NEUROLOGICAL DISORDERS, T. N. I. of et S. rt PA STROKE STUDY GROUP (1995). « Tissue Plasminogen Activator for Acute Ischemic Stroke ». Dans : *New England Journal of Medicine* 333.24, p. 1581–1588. URL : <http://dx.doi.org/10.1056/NEJM199512143332401>.

- NGUYEN, V. H., G. COOPERMAN, N. MENENZES, C. J. LOPEZ, C. MELINOSKY, O. WU, H. AY, Y. LIU, J. NUUTINEN, H. J. ARONEN, J. KARONEN, A. G. SORENSEN, W. J. KOROSHETZ et H. H. PIEN (2008). « Stroke Tissue Outcome Prediction Using A Spatially-Correlated Model ». Dans : *Pan Pacific Imaging conference*. URL : http://www.researchgate.net/profile/Walter_Koroshetz/publication/228805650_Stroke_Tissue_Outcome_Prediction_Using_A_Spatially-Correlated_Model/links/02bfe513dcf1338d09000000.pdf.
- OLIVOT, J.-M., M. MLYNASH, V. N. THIJS, S. KEMP, M. G. LANSBERG, L. WECHSLER, R. BAMMER, M. P. MARKS et G. W. ALBERS (2009a). « Optimal Tmax Threshold for Predicting Penumbra Tissue in Acute Stroke ». Dans : *Stroke* 40.2, p. 469–475. URL : <http://stroke.ahajournals.org/content/40/2/469.abstract>.
- OLIVOT, J.-M., M. MLYNASH, V. N. THIJS, A. PURUSHOTHAM, S. KEMP, M. G. LANSBERG, L. WECHSLER, R. BAMMER, M. P. MARKS et G. W. ALBERS (2009b). « Relationships Between Cerebral Perfusion and Reversibility of Acute Diffusion Lesions in DEFUSE : Insights from RADAR ». Dans : *Stroke* 40.5, p. 1692–1697. URL : <http://dx.doi.org/10.1161/STROKEAHA.108.538082>.
- OSHER, S. et R. FEDKIW (2003). *Level Set Methods and Dynamic Implicit Surfaces*. Springer. URL : <http://www.springer.com/us/book/9780387954820>.
- ØSTERGAARD, L., K. Y. JÓNSDÓTTIR et K. MOURIDSEN (2009). « Predicting tissue outcome in stroke : new approaches ». Dans : *Current Opinion in Neurology* 22.1, 54—59. URL : <http://dx.doi.org/10.1097/wco.0b013e328320d297>.
- ØSTERGAARD, L. C., I. N. de la Santé et de RECHERCHE MEDICALE, U. C. BERNARD, F. P. I. d'Investigació Biomèdica de GIRONA, U. of CAMBRIDGE, U. HAMBURG-EPPENDORF, S. S. ENGINEERING et DIMAC. (2006). « I-know Integration Information from Molecule to Man ». Dans : *European Commision's Sixth Freamework Programme*. URL : http://vph-portal.eu/documents/10180/35174/I_Know_factsheet.pdf.
- OZENNE, B., T.-H. CHO, I. K. MIKKELSEN, M. HERMIER, L. RIBE, G. THOMALLA, S. PEDRAZA, J.-C. BARON, P. ROY, Y. BERTHEZENE, N. NIGHOGHOSSIAN, L. ØSTERGAARD et D. MAUCORT-BOULCH (2015). « Evaluation of Early Reperfusion Criteria in Acute Ischemic Stroke ». Dans : *Journal of Neuroimaging*. URL : <http://dx.doi.org/10.1111/jon.12255>.
- PARSONS, M. W., Q. YANG, P. A. BARBER, D. G. DARBY, P. M. DESMOND, R. P. GERRATY, B. M. TRESS et S. M. DAVIS (2001). « Perfusion Magnetic Resonance Imaging Maps in Hyperacute Stroke : Relative Cerebral Blood Flow Most Accurately Identifies Tissue Destined to Infarct ». Dans : *Stroke* 32.7, p. 1581–1587. URL : <http://dx.doi.org/10.1161/01.STR.32.7.1581>.
- PEREYRA, M., N. DOBIGEON, H. BATATIA et J. Y. TOURNERET (2013). « Estimating the granularity coefficient of a Potts-Markov random field within an MCMC algorithm ». Dans : *IEEE Transactions on Image Processing* 22.6, p. 2385–2397. URL : <http://dx.doi.org/10.1109/tip.2013.2249076>.
- PRAKASH, K. B., V. GUPTA, M. BILELLO, N. J. BEAUCHAMP et W. L. NOWINSKI (2006). « Identification, Segmentation, and Image Property Study of Acute Infarcts in Diffusion-Weighted Images by Using a Probabilistic Neural Network and Adaptive Gaussian Mixture Model ». Dans : *Academic Radiology* 13.12, p. 1474–1484. URL : <http://dx.doi.org/10.1016/j.acra.2006.09.045>.

- RAMSAY, J. et B. W. SILVERMAN (2005). *Functional Data Analysis*. Springer. URL : <http://www.springer.com/us/book/9780387400808>.
- REKIK, I., S. ALLASSONNIÈRE, T. K. CARPENTER et J. M. WARDLAW (2012). « Medical image analysis methods in MR/CT-imaged acute-subacute ischemic stroke lesion : Segmentation, prediction and insights into dynamic evolution simulation models. A critical appraisal ». Dans : *NeuroImage : Clinical* 1.1, p. 164–178. URL : <http://dx.doi.org/10.1016/j.nicl.2012.10.003>.
- REKIK, I., S. ALLASSONNIÈRE, T. K. CARPENTER et J. M. WARDLAW (2013). « Spatiotemporal Dynamic Simulation of Acute Perfusion/Diffusion Ischemic Stroke Lesions Evolution : A Pilot Study Derived from Longitudinal MR Patient Data ». Dans : *Computational and Mathematical Methods in Medicine*. URL : <http://dx.doi.org/10.1155/2013/283593>.
- (2014). « Using longitudinal metamorphosis to examine ischemic stroke lesion dynamics on perfusion-weighted images and in relation to final outcome on T2-w images ». Dans : *Neuroimage : Clinical* 5, 332–340. URL : <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4141979/>.
- REVOL, C. et M. JOURLIN (1997). « A new minimum variance region growing algorithm for image segmentation ». Dans : *Pattern Recognition Letters* 18.3, p. 249–258. URL : [http://dx.doi.org/10.1016/S0167-8655\(97\)00012-3](http://dx.doi.org/10.1016/S0167-8655(97)00012-3).
- REVOL-MULLER, C., F. PEYRIN, Y. CARRILLON et C. ODET (2002). « Automated 3D region growing algorithm based on an assessment function ». Dans : *Pattern Recognition Letters* 23.1–3, p. 137–150. URL : [http://dx.doi.org/10.1016/S0167-8655\(01\)00116-7](http://dx.doi.org/10.1016/S0167-8655(01)00116-7).
- RINGER, T. M., T. NEUMANN-HAEFELIN, R. A. SOBEL, M. E. MOSELEY et M. A. YENARI (2001). « Reversal of Early Diffusion-Weighted Magnetic Resonance Imaging Abnormalities Does Not Necessarily Reflect Tissue Salvage in Experimental Cerebral Ischemia ». Dans : *Stroke* 32.10, p. 2362–2369. URL : <http://dx.doi.org/10.1161/hs1001.096058>.
- RIVERS, C., J. WARDLAW, P. ARMITAGE, M. BASTIN, T. CARPENTER, V. CVORO, P. HAND et M. DENNIS (2006). « Do Acute Diffusion- and Perfusion-Weighted MRI Lesions Identify Final Infarct Volume in Ischemic Stroke ? » Dans : *Stroke* 37.1, p. 98–104. URL : <http://dx.doi.org/10.1161/01.STR.0000195197.66606.bb>.
- ROSE, J.-L. (2008). « Croissance de région variationnelle et contraintes géométriques tridimensionnelles pour la segmentation d'image ». Dans : *Thèse*. URL : <http://theses.insa-lyon.fr/publication/2008ISAL0108/these.pdf>.
- ROSEN, B., J. BELLIVEAU, J. VEVEA et T. BRADY (1990). « Perfusion imaging with NMR Contrast Agents ». Dans : *Magnetic Resonance in Medicine* 14.1, p. 249–265. URL : <http://dx.doi.org/10.1002/mrm.1910140211>.
- ROSSO, C., N. HEVIA-MONTIEL, S. DELTOUR, E. BARDINET, D. DORMONT, S. CROZIER, S. BAILLET et Y. SAMSON (2009). « Prediction of Infarct Growth Based on Apparent Diffusion Coefficients : Penumbral Assessment without Intravenous Contrast Material ». Dans : *Radiology* 250.1, p. 184–192. URL : <http://dx.doi.org/10.1148/radiol.2493080107>.

- SAAD, N. M., S. ABU-BAKAR, S. MUDA, M. MOKJI et A. ABDULLAH (2012). « Fully Automated Region Growing Segmentation of Brain Lesion in Diffusion-weighted MRI ». Dans : *International Journal of Computer Science* 39.2, p. 1–10. URL : http://www.iaeng.org/IJCS/issues_v39/issue_2/IJCS_39_2_03.pdf.
- SAVER, J. L. (2006). « Time Is Brain—Quantified ». Dans : *Stroke* 37.1, p. 263–266. URL : <http://dx.doi.org/10.1161/01.STR.0000196957.55928.ab>.
- SCALZO, F., Q. HAO, J. R. ALGER, X. HU et D. S. LIEBESKIND (2012). « Regional Prediction of Tissue Fate in Acute Ischemic Stroke ». Dans : *Ann Biomed Eng* 40.10, p. 2177–2187. URL : <http://dx.doi.org/10.5772/22890>.
- SCHAEFER, P. W., P. E. GRANT et R. G. GONZALEZ (2000). « Diffusion-weighted MR Imaging of the Brain ». Dans : *Radiology* 217.2, p. 331–345. URL : <http://dx.doi.org/10.1007/b137507>.
- SCHELLINGER, P., R. BRYAN, L. CAPLAN, J. DETRE, R. EDELMAN, C. JAIGOBIN, C. KIDWELL, J. MOHR, M. SLOAN, A. SORENSEN et S. WARACH (2010). « Evidence-based guideline : The role of diffusion and perfusion MRI for the diagnosis of acute ischemic stroke ». Dans : *Neurology* 75.1, 177–185. URL : <http://dx.doi.org/10.1212/wnl.0b013e3181f367ae>.
- SEGHER, M. L., A. RAMLACKHANSINGH, J. CRINION, A. P. LEFF et C. J. PRICE (2008). « Lesion identification using unified segmentation-normalisation models and fuzzy clustering ». Dans : *NeuroImage* 41.4, p. 1253–1266. URL : <http://dx.doi.org/10.1016/j.neuroimage.2008.03.028>.
- SHEN, Q. et T. Q. DUONG (2008a). « Quantitative prediction of ischemic stroke tissue fate ». Dans : *NMR in Biomedicine* 21.8, p. 839–848. URL : <http://dx.doi.org/10.1002/nbm.1264>.
- SHEN, S., A. SZAMEITAT et A. STERR (2008b). « Detection of Infarct Lesions From Single MRI Modality Using Inconsistency Between Voxel Intensity and Spatial Location ». Dans : *IEEE Transactions on Information Technology in Biomedicine* 12.4, p. 532–540. URL : <http://dx.doi.org/10.1109/TITB.2007.911310>.
- SHEN, S., A. J. SZAMEITAT et A. STERR (2010). « An improved lesion detection approach based on similarity measurement between fuzzy intensity segmentation and spatial probability maps ». Dans : *Magnetic Resonance Imaging* 28.2, p. 245–254. URL : <http://www.sciencedirect.com/science/article/pii/S0730725X09001751>.
- SHIH, L. C., J. L. SAVER, J. R. ALGER, S. STARKMAN, M. C. LEARY, F. VINUELA, G. DUCKWILER, Y. P. GOBIN, R. JAHAN, J. P. VILLABLANCA, P. M. VESPA et C. S. KIDWELL (2003). « Perfusion-Weighted Magnetic Resonance Imaging Thresholds Identifying Core, Irreversibly Infarcted Tissue ». Dans : *Stroke* 34.1, p. 1425–1430. URL : <http://dx.doi.org/10.1161/01.str.0000072998.70087.e9>.
- SOLTANIAN-ZADEH, H., H. BAGHER-EBADIAN, J. R. EWING, P. D. MITSIAS, A. KAPKE, M. LU, Q. JIANG, S. C. PATEL et M. CHOPP (2007). « Multiparametric Iterative Self-Organizing Data Analysis of Ischemic Lesions Using Pre- or Post-Gd T1 MRI ». Dans : *Cerebrovascular Diseases* 23.2-3, p. 91–102. URL : <http://dx.doi.org/10.1159/000097044>.
- SOLTANIAN-ZADEH, H., M. PASNOOR, R. HAMMOUD, M. A. JACOBS, S. C. PATEL, P. D. MITSIAS, R. A. KNIGHT, Z. G. ZHENG, M. LU et M. CHOPP (2003). « MRI tissue characterization of experimental cerebral ischemia in rat ». Dans : *Journal of Magnetic Resonance Imaging* 17.4, p. 398–409. URL : <http://dx.doi.org/10.1002/jmri.10256>.

- STEIN, B., D. LISIN, J. HOROWITZ, E. RISEMAN et G. WHITTEN (2001). « Statistical and Deformable Model Approaches to the Segmentation of MR Imagery and Volume Estimation of Stroke Lesions ». Dans : *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2001*. Lecture Notes in Computer Science 2208. Sous la dir. de W. NIESSEN et M. VIERGEVER, p. 829–836. URL : http://dx.doi.org/10.1007/3-540-45468-3_99.
- STORATH, M. et A. WEINMANN (2014). « Fast partitioning of vector-valued images ». Dans : *SIAM Journal on Imaging Sciences* 7.3, p. 1826–1852. URL : <http://dx.doi.org/10.1137/130950367>.
- SWEENEY, E. M., R. T. SHINOHARA, N. SHIEE, F. J. MATEEN, A. A. CHUDGAR, J. L. CUZZOCREO, P. A. CALABRESI, D. L. PHAM, D. S. REICH et C. M. CRAINICEANU (2013). « OASIS is Automated Statistical Inference for Segmentation, with applications to multiple sclerosis lesion segmentation in MRI. » Dans : *NeuroImage : Clinical* 2, p. 402–13. URL : <http://dx.doi.org/10.1016/j.nicl.2013.03.002>.
- SWEENEY, E. M., J. T. VOGELSTEIN, J. L. CUZZOCREO, P. A. CALABRESI, D. S. REICH, C. M. CRAINICEANU et R. T. SHINOHARA (avr. 2014). « A Comparison of Supervised Machine Learning Algorithms and Feature Vectors for MS Lesion Segmentation Using Multimodal Structural MRI ». Dans : *PLoS ONE* 9.4. URL : <http://dx.doi.org/10.1371/journal.pone.0095753>.
- TAHA, A. A. et A. HANBURY (2015). « Metrics for evaluating 3D medical image segmentation : analysis, selection, and tool ». Dans : *BMC Medical Imaging* 15.1. URL : <http://dx.doi.org/10.1186/s12880-015-0068-x>.
- THE ATLANTIS, ECASS, AND NINDS RT-PA STUDY GROUP INVESTIGATORS (2004). « Association of outcome with early stroke treatment : pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials ». Dans : *The Lancet* 363.1, p. 768–774. URL : [http://dx.doi.org/10.1016/S0140-6736\(04\)15692-4](http://dx.doi.org/10.1016/S0140-6736(04)15692-4).
- THIJS, V., A. ADAMI, T. NEUMANN-HAEFELIN, M. MOSELEY, M. MARKS et A. G.W. (2001). « Relationship between severity of MR perfusion deficit and DWI lesion evolution ». Dans : *Neurology* 57.7, p. 1205–11. URL : <http://dx.doi.org/10.1212/wnl.57.7.1205>.
- THIRION, B., G. FLANDIN, P. PINEL, A. ROCHE, P. CIUCIU et J.-B. POLINE (2006). « Dealing with the shortcomings of spatial normalization : Multi-subject parcellation of fMRI datasets ». Dans : *Human Brain Mapping* 27.8, p. 678–693. URL : <http://dx.doi.org/10.1002/hbm.20210>.
- TJELMELAND, H. et J. BESAG (1998). « Markov Random Fields with Higher-order Interactions ». Dans : *Scandinavian Journal of Statistics* 25.3, p. 415–433. URL : <http://dx.doi.org/10.1111/1467-9469.00113>.
- TOMASI, C. et R. MANDUCHI (1998). « Bilateral filtering for gray and color images ». Dans : *Sixth International Conference on Computer Vision*, p. 839–846. URL : <http://dx.doi.org/10.1109/ICCV.1998.710815>.
- USINSKAS, A., E. PRANCKEVICIENE, T. WITTENBERG, P. HASTREITER et B. F. TOMANDL (2003). « Automatic Ischemic Stroke Segmentation Using Various Techniques ». Dans : *Neural Networks and Soft Computing* 19, p. 498–503. URL : http://dx.doi.org/10.1007/978-3-7908-1902-1_76.
- USINSKAS, A., R. A. DOBROVOLSIS et B. F. TOMANDL (2004). « Ischemic Stroke Segmentation on CT Images Using Joint Features ». Dans : *Informatica* 15.2, p. 283–290. URL : <http://dl.acm.org/citation.cfm?id=1413509.1413519>.

- VACHIER, C. et F. MEYER (2005). « The Viscous Watershed Transform ». Dans : *Journal of Mathematical Imaging and Vision* 22.2-3, p. 251–267. URL : <http://dx.doi.org/10.1007/s10851-005-4893-3>.
- VAN CALSTER, B., A. J. VICKERS, M. J. PENCINA, S. G. BAKER, D. TIMMERMAN et E. W. STEYERBERG (2013). « Evaluation of Markers and Risk Prediction Models : Overview of Relationships between NRI and Decision-Analytic Measures ». Dans : *Medical Decision Making* 33.4, p. 490–501. URL : <http://dx.doi.org/10.1177/0272989X12470757>.
- VERMEER, S. E., M. HOLLANDER, E. J. van DIJK, A. HOFMAN, P. J. KOUDSTAAL et M. M. BRETELER (2003). « Silent Brain Infarcts and White Matter Lesions Increase Stroke Risk in the General Population : The Rotterdam Scan Study ». Dans : *Stroke* 34.5, p. 1126–1129. URL : <http://dx.doi.org/10.1161/01.str.0000068408.82115.d2>.
- VERNOOIJ, M. W., M. A. IKRAM, H. L. TANGHE, A. J. VINCENT, A. HOFMAN, G. P. KRESTIN, W. J. NIESSEN, M. M. BRETELER et A. van der LUGT (2007). « Incidental Findings on Brain MRI in the General Population ». Dans : *New England Journal of Medicine* 357.18, p. 1821–1828. URL : <http://dx.doi.org/10.1056/NEJMoA070972>.
- WALTER, S. D. (2005). « The partial area under the summary ROC curve ». Dans : *Statistics in Medicine* 24.13, p. 2025–2040. URL : <http://dx.doi.org/10.1002/sim.2103>.
- WEINMAN, J., G. BISSIAS, J. HOROWITZ, E. RISEMAN et A. HANSON (2003). « Nonlinear Diffusion Scale-Space and Fast Marching Level Sets for Segmentation of MR Imagery and Volume Estimation of Stroke Lesions ». Dans : *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003*. Lecture Notes in Computer Science 2879. Sous la dir. de R. ELLIS et T. PETERS, p. 496–504. URL : http://dx.doi.org/10.1007/978-3-540-39903-2_61.
- WILKE, M., B. de HAAN, H. JUENGER et H.-O. KARNATH (2011). « Manual, semi-automated, and automated delineation of chronic brain lesions : A comparison of methods ». Dans : *NeuroImage* 56.4, p. 2038–2046. URL : <http://dx.doi.org/10.1016/j.neuroimage.2011.04.014>.
- WOOD, S. N. (2003). « Thin plate regression splines ». Dans : *Journal of the Royal Statistic Society* 65.1, p. 95–114. URL : <http://dx.doi.org/10.1111/1467-9868.00374>.
- WOOD, S. N. (2006). *Generalized Additive Models : An Introduction with R*. Texts in Statistical Science Series. Chapman et Hal/CRC. URL : <https://www.crcpress.com/Generalized-Additive-Models-An-Introduction-with-R/Wood/9781584884743>.
- WOOLRICH, M. W., T. E. J. BEHRENS, C. F. BECKMANN et S. M. SMITH (2005). « Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data ». Dans : *IEEE Transactions on Medical Imaging* 24.1, p. 1–11. URL : <http://dx.doi.org/10.1109/TMI.2004.836545>.
- WU, O., S. CHRISTENSEN, P. ROSA-NETO, N. HJORT, A. RODELL, R. M. DIJKHUIZEN, J. FIEHLER, J. RÖTHER et L. ØSTERGAARD (2004). « Anatomy as a parameter in multiparametric MRI-based predictive algorithms ». Dans : *Proceedings of the International Society for Magnetic Resonance in Medicine*. URL : <http://cds.ismrm.org/ismrm-2004/Files/001392.pdf>.

- WU, O., L. SCHWAMM et A. SORESENSEN (2011). « Imaging stroke patients with unclear onset times ». Dans : *Neuroimaging clinics of North America* 21.2, p. 327–44. URL : <http://dx.doi.org/10.1016/j.nic.2011.02.008>.
- WU, O., W. J. KOROSHETZ, L. ØSTERGAARD, F. S. BUONANNO, W. A. COPEN, R. G. GONZALEZ, G. RORDORF, B. R. ROSEN, L. H. SCHWAMM, R. M. WEISSKOFF et A. G. SORESENSEN (2001). « Predicting Tissue Outcome in Acute Human Cerebral Ischemia Using Combined Diffusion- and Perfusion-Weighted MR Imaging ». Dans : *Stroke* 32.4, p. 933–942. URL : <http://dx.doi.org/10.1161/01.STR.32.4.933>.
- WU, O., L. ØSTERGAARD, R. M. WEISSKOFF, T. BENNER, B. R. ROSEN et A. G. SORESENSEN (2003). « Tracer arrival timing-insensitive technique for estimating flow in MR perfusion-weighted imaging using singular value decomposition with a block-circulant deconvolution matrix ». Dans : *Magnetic Resonance in Medicine* 50.1, p. 164–174. URL : <http://dx.doi.org/10.1002/mrm.10522>.
- WU, O., S. CHRISTENSEN, N. HJORT, R. M. DIJKHUIZEN, T. KUCINSKI, J. FIEHLER, J. THOMALLA Götz and Røther et L. ØSTERGAARD (2006). « Characterizing physiological heterogeneity of infarction risk in acute human ischaemic stroke using MRI ». Dans : *Brain* 129.9, p. 2384–2393. URL : <http://dx.doi.org/10.1093/brain/awl183>.
- WU, O., R. M. DIJKHUIZEN et A. G. SORESENSEN (2010). « Multiparametric Magnetic Resonance Imaging of Brain Disorders ». Dans : *Topics in Magnetic Resonance Imaging* 21.2, p. 129–138. URL : <http://dx.doi.org/10.1097/rmr.0b013e31821e56c2>.
- WU, O., T. BENNER, L. ROCCATAGLIATA, M. ZHU, P. SCHAEFER, A. SORESENSEN et A. SINGHAL (2012). « Evaluating effects of normobaric oxygen therapy in acute stroke with MRI-based predictive models ». Dans : *Medical Gas Research* 2.1, p. 5. URL : <http://dx.doi.org/10.1186/2045-9912-2-5>.
- YOSHITA, M., E. FLETCHER et C. DECARLI (2005). « Current Concepts of Analysis of Cerebral White Matter Hyperintensities on Magnetic Resonance Imaging ». Dans : *Topics in Magnetic Resonance Imaging* 16.6, p. 399–407. URL : <http://dx.doi.org/10.1097/01.rmr.0000245456.98029.a8>.
- ZAHARCHUK, G., R. BAMMER, M. STRAKA, R. D. NEWBOULD, J. ROSENBERG, J.-M. OLIVOT, M. MLYNASH, M. G. LANSBERG, N. E. SCHWARTZ, M. M. MARKS, G. W. ALBERS et M. E. MOSELEY (2009). « Improving dynamic susceptibility contrast MRI measurement of quantitative cerebral blood flow using corrections for partial volume and nonlinear contrast relaxivity : A xenon computed tomographic comparative study ». Dans : *Journal of Magnetic Resonance Imaging* 30.4, p. 743–752. URL : <http://dx.doi.org/10.1002/jmri.21908>.
- ZHANG, H., N. LU, C. FENG, S. THURSTON, Y. XIA et X. TU (2011). « On Fitting Generalized Linear Mixed-effects Models for Binary Responses using Different Statistical Packages ». Dans : *Statistics in medicine* 30.20, p. 2562–2572. URL : <http://dx.doi.org/10.1002/sim.4265>.
- ZHANG, X., T. D. JOHNSON, R. J. A. LITTLE et Y. CAO (2008). « Quantitative magnetic resonance image analysis via the EM algorithm with stochastic variation ». Dans : *The Annals of Applied Statistics* 2.2, p. 736–755. URL : <http://dx.doi.org/10.1214/07-aos157>.
- ZHANG, Y., M. BRADY et S. SMITH (2001). « Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm ». Dans : *IEEE Transactions on Medical Imaging* 20.1, p. 45–57. URL : <http://dx.doi.org/10.1109/42.906424>.

ZOPPO, G. J. del et J. M. HALLENBECK (2000). « Advances in the Vascular Pathophysiology of Ischemic Stroke ». Dans : *Thrombosis Research* 98.3, p. 73 –81. URL : [http://dx.doi.org/10.1016/s0049-3848\(00\)00218-8](http://dx.doi.org/10.1016/s0049-3848(00)00218-8).

Annexe A: Présentation de la cohorte de patients étudiée

1 Critères d'inclusion des patients

Les critères d'inclusion des patients dans l'étude I-know sont :

- Score NIHSS d'au moins 4 points
- Imagerie en diffusion-perfusion en accord avec la présence d'un AVC ischémique de la circulation antérieure
- Délai entre le début des symptômes et l'examen IRM de moins 6 heures si le patient est thrombolysé, sinon de moins de 12 heures.

Ces critères ont pour but d'homogénéiser les AVC recueillis en excluant les très petits AVC, les accidents ischémiques transitoires et les AVC hémorragiques. Les patients pour lesquels le délai depuis le début des symptômes est inconnu sont également exclus de l'étude.

2 Schéma de l'étude

L'étude I-know prévoit quatre temps de mesure IRM :

- **à l'admission du patient (H0)** : acquisition des paramètres de diffusion, perfusion, du T1, du T2 et du T2 FLAIR. La lésion est contournée par le médecin. Identification du type d'AVC et du lieu d'occlusion.
- **quelques heures après l'admission du patient (H2)** : acquisition des paramètres de diffusion, perfusion, du T2 et du T2 FLAIR. La lésion est contournée par le médecin.
- **quelques jours après l'admission du patient (J2)** : acquisition des paramètres de diffusion, du T2 et du T2 FLAIR.
- **à stabilisation de l'AVC (M1)** : acquisition des paramètres de diffusion, du T2 et du T2 FLAIR. La lésion est contournée par le médecin.

Les données biologiques (prélèvement sanguin) sont recueillies à l'admission du patient. Des données cliniques sont également recueillies (historique médical). En particulier le score de handicap NIHSS est calculé à l'admission du patient (H0), à H2, ainsi qu'à stabilisation de l'AVC (M1). Un score alternatif, le score mRS (échelle de Rankin modifiée ou modified Rankin scale en anglais), est aussi calculé à M1.

3 Données disponibles

172 patients ont été inclus dans l'étude. Nous avons eu accès aux données de 63 patients pour lesquels la quasi-totalité de l'information prévue était récoltée : il manquait 31 paramètres IRM (sur les 1827 à récolter) dont 13 pour le même patient (qui a été exclu de l'ensemble des analyses). Les données cliniques et biologiques étaient complètes excepté pour un patient.

Le tableau A.1 synthétise la distribution des temps de prise en charge et de mesure IRM pour les différents patients. Le délai entre la première IRM et le début des symptômes varie entre 50 minutes et 6h 43 minutes. Ainsi la plupart des patients ont pu être thrombolysés (68 %). Le second examen IRM, inhabituel par rapport aux protocoles cliniques usuels, a pour but d'identifier des reperfusion précoces et doit donc être fait rapidement après que le traitement ait agit. Compte tenu des contraintes pratiques, le délai initialement fixé à deux heures a été bien respecté avec une médiane à 2 heures 50 minutes. L'instant de réalisation de l'IRM finale est très variable (entre une semaine et trois mois). Les principales évolutions de l'AVC ayant lieu dans les dix premières heures (SAVER, 2006), l'AVC s'est généralement stabilisé au bout d'une semaine. Aussi cette variabilité devrait peu impacter le résultat des acquisitions IRM.

Délai entre	minimum	1 ^{ier} quartile	médiane	3 ^{ième} quartile	maximum
Symptômes - Hôpital	23 min	1h 15 min	1h 45 min	2h 40 min	5h 30 min
Hôpital - IRM (H0)	0 min	10 min	22 min	31 min	2h min
IRM (H0) - Traitement	10 min	34 min	54 min	1h	1h 22 min
IRM (H0) - IRM (H2)	1h 33 min	2h 28 min	2h 50 min	3h 11	4h 52 min
IRM (H2) - IRM (M1)	7 jours	17 jours	30 jours	89 jours	109 jours

TABLE A.1 – Instants de prise en charge et de mesure IRM selon les patients.

Les principales variables descriptives de la cohorte sont reportées dans le tableau A.2. La cohorte est principalement composée de patients âgés (médiane à 70 ans) ayant des AVC de taille très variable (de quelques mL à plus 100 mL). Globalement le score NIHSS diminue suite à la prise en charge à l'hôpital. Pour une partie des patients, des événements hémorragiques mineurs ont été reportés (HI 1 ou 2). Aucun patient n'a été affecté par une hémorragie significative (PH 1 ou 2). Enfin une recanalisation (dissolution du matériel obstruant l'artère) partielle ou complète est observée à deux jours pour près de la moitié des patients.

Variable continue	minimum	1 ^{er} quartile	médiane	3 ^{ème} quartile	maximum
Âge	30 ans	62 ans	69 ans	77 ans	86 ans
Score NIHSS à H0	1	6	10	16	24
Score NIHSS à H2	0	4	7	14	23
Volume de lésion à H0 (mL)	0	3,6	10,7	25,6	175,3
Score NIHSS final	0	0	2	5,5	21
Volume de lésion à final (mL)	0	2,7	9,8	40,0	238,5
Variable catégorielle	Type 1	Type 2	Type 3	Type 4	Type 5
Sexe	Homme (60%)	Femme (40%)			
Antécédent d'AVC	Oui (11%)	Non (89%)			
Facteur(s) de risque	Hypertension (63%)	Diabète (6%)	Fumeur (23 %)	FA (16 %)	Artériopathie (11%)
Hôpital	Aarhus (29%)	Gérone (3%)	Hambourg (10%)	Lyon (59%)	
Éthiologie	Embole (40%)	Sténose (16%)	Dissection (5%)	Indéterminé (15%)	Autres (24%)
Médication(s)	Thrombolyse (69%)	Statine (23%)	Antihypertenseur (55%)	Aspirine (35%)	
Hémorragie (H2)	Non (90%)	HI1 (8%)	HI2 (2%)	PH1 (0%)	PH2 (0%)
Hémorragie (D2)	Non (61%)	HI1 (26%)	HI2 (13%)	PH1 (0%)	PH2 (0%)
Recanalisation (H2)	Aucun (63%)	Faible (3%)	Partielle (22%)	Complète (12%)	
Recanalisation (D2)	Aucune (37%)	Faible (5%)	Partielle (35%)	Complète (23%)	
Score mRS (M1)	0 (20%)	1 (25%)	2 (18%)	3 (16%)	≥ 4 (21%)

TABLE A.2 – Variables descriptives de la cohorte étudiée. FA : antécédent de fibrillation auriculaire. HI et PH correspondent à une échelle de classification des hémorragies décrite dans (LARRUE et al., 1997)

Annexe B: Présentation des méthodes de prédiction

Cette annexe présente les approches prédictives utilisées dans la section 3. Les quatre premières approches sont des méthodes de modélisation statistique pour données binaires : régression logistique et ses extensions pour données hiérarchiques, spatialement corrélées ou pour des facteurs de risque non linéaires. Ensuite, une introduction aux arbres de classification CART précède la présentation de l'algorithme Adaboost puisque ce dernier est composé d'une combinaison d'arbres CART.

1 La régression logistique (LR)

Comme l'ensemble des modèles de régression, la régression logistique va chercher à estimer la distribution de probabilité ayant généré les données. Dans le cas de données binaires, on suppose que chaque observation y_i est issue d'une loi de Bernoulli. Il faut donc déterminer l'espérance $p_i = \mathbb{P}[y_i = 1]$. Pour cela l'espérance est reliée aux variables prédictives à l'aide d'une fonction de lien η . Comme son nom l'indique, la régression logistique suppose une fonction de lien *logit* et suppose un effet linéaire des variables prédictives sur l'échelle logistique. On aboutit donc à la distribution de probabilité suivante :

$$y_i \sim \mathcal{B}(p_i) \text{ avec } \text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \alpha + \sum_{k=1}^K \mathcal{X}_{i,k} \beta_k \quad (\text{B.1})$$

L'espérance est le résultat de la modélisation des observations : elle sera la variable latente Y^* qui servira de biomarqueur pour le risque de nécrose. En notant $\boldsymbol{\beta} = (\beta_k)_{k \in \{1, \dots, K\}}$ avec K le nombre de variables prédictives, si l'on connaît les paramètres $(\alpha, \boldsymbol{\beta})$ du modèle, on peut calculer le risque de nécrose pour une nouvelle observation y_0 à partir des mesures \mathcal{X}_0 :

$$p_0 = \mathbb{P}[y_0 = 1 | \mathcal{X}_0, \alpha, \boldsymbol{\beta}] = \frac{1}{1 + \exp(-\alpha - \sum_{k=1}^K \mathcal{X}_{0,k} \beta_k)}$$

Ainsi une variable prédictive avec un coefficient estimé β positif sera considérée comme un facteur de risque. Un coefficient négatif correspond à une variable protectrice et un coefficient nul à l'absence d'influence de la variable. Le paramètre α permet essentiellement de calibrer le modèle en fonction de la prévalence de l'évènement.


L'estimation du modèle logistique se fait par maximum de vraisemblance ce qui confère un certain nombre de propriétés à l'estimation : estimateur convergent, asymptotiquement efficace et normalement distribué (voir par exemple la section 2.4 de l'ouvrage de Wood - S. N. WOOD, 2006). Cela autorise aussi l'utilisation des tests statistiques usuels pour évaluer la pertinence du modèle et des variables prédictives. Afin d'effectuer la maximisation on suppose que les observations sont indépendantes, ce qui permet de se ramener au problème suivant :

Chercher α et β maximisant

$$\begin{aligned} L_v(\alpha, \beta | x, y) &= \mathbb{P}[Y = y | \alpha, \beta, \mathcal{X} = x] \\ &= \prod_{i=1}^n \mathbb{P}[Y_i = y_i | \alpha, \beta, \mathcal{X}_i = x_i]^{y_i} * (1 - \mathbb{P}[Y_i = y_i | \alpha, \beta, \mathcal{X}_i = x_i])^{1-y_i} \end{aligned}$$

où $\mathbb{P}[Y = y_i | \alpha, \beta, \mathcal{X}_i = x_i] = p_i$ est défini par l'équation B.1. Cette maximisation est équivalente à résoudre un problème de moindres carrés pondérés non linéaires (section 2.1 de l'ouvrage de Wood - S. N. WOOD, 2006) :

$$\text{Chercher } \alpha, \beta \text{ minimisant : } \|\sqrt{\text{Var}[\mathbf{p}]}^{-1}(\mathbf{y} - \mathbf{p})\|^2 \quad (\text{B.2})$$

où la dépendance de $\mathbf{p} = (p_i)_{i \in \{1, \dots, n\}}$ en α et β est non-linéaire (équation B.1). La résolution de ce problème se fait en utilisant l'algorithme IRLS (moindres carrés pondérés itérés). En pratique, la fonction `glm` de la librairie *stats* du logiciel  a été utilisée pour estimer ce modèle.

Le modèle logistique repose sur un certain nombre d'hypothèses :

- indépendance des observations
- linéarité des risques associés à chaque variable prédictive
- absence d'interaction entre les variables prédictives
- lien logit entre les variables prédictives et l'espérance

qui peuvent mettre en défaut la qualité de l'ajustement du modèle et donc des prédictions.

2 La régression logistique additive (ALR)

La régression logistique additive est une généralisation de la régression logistique permettant une modélisation non linéaire du risque associé à chaque variable. Une fonction de risque, notée f_k , est alors associée à la $k^{\text{ième}}$ variable prédictive :

$$\text{logit}(p_i) = \alpha + \sum_{k=1}^K f_k(\mathcal{X}_{i,k})$$

Si f_k est une fonction linéaire, on se ramène à la régression logistique usuelle. Ce modèle est très flexible puisqu'aucune contrainte n'est imposée aux fonctions f_k . Il généralise donc l'utilisation de polynômes (i.e. $\beta_k \mathcal{X}_{i,k} + \gamma_k \mathcal{X}_{i,k}^2 + \dots$) qui nécessitent d'explicitement la forme de la

fonction de risque et limitent donc la souplesse du modèle. De plus il est possible de considérer des interactions entre variables en utilisant des fonctions multivariées du type $h_{kl}(\mathcal{X}_k, \mathcal{X}_l)$.

L'estimation du modèle, c'est-à-dire l'estimation des fonctions de risque f_k et du paramètre α , repose sur la redéfinition du critère d'adéquation aux données. En effet, d'une part, la flexibilité induite par les fonctions f_k augmente le risque de surapprentissage, et d'autre part, le nombre de paramètres à estimer pour identifier les fonctions f_k pouvant être grand, le problème peut devenir mal posé (i.e. nombre de paramètres supérieur au nombre de degrés de liberté). La solution consiste alors à supposer les fonctions f_k régulières : c'est-à-dire qu'une petite variation de la valeur des variables prédictives n'entraîne pas une grande variation du risque associé. Cette hypothèse est souvent très raisonnable (voire souhaitable) lorsque l'on étudie des phénomènes biologiques. Pour tenir compte de cette hypothèse, la vraisemblance va être pénalisée par la tortuosité des fonctions f_k : on souhaite optimiser un compromis entre l'adéquation aux données et la régularité des fonctions de risque. Ainsi le problème des moindres carrés (B.2) se réécrit :

$$\text{Chercher } \alpha, f_1 \dots f_k \text{ minimisant : } \|\sqrt{\text{Var}[\mathbf{p}]}^{-1}(\mathbf{y} - \mathbf{p})\|^2 + \sum_{k=1}^K \lambda_k \int_{-\infty}^{\infty} f_k''(x)^2 dx \quad (\text{B.3})$$

où la dérivée seconde (ou courbure) de la fonction a été utilisée pour mesurer sa tortuosité. Les paramètres λ_k sont des paramètres qui règlent pour chaque fonction de risque le compromis adéquation/régularité. Les splines plaques minces fournissent la forme générale des fonctions de risque solutions de ce problème :

$$\widehat{f}_k(x) = \sum_{j=1}^m \gamma_{j,k} \phi_j(x) + \sum_{i=1}^n \delta_{i,k} \eta_{m,1}(\|x - x_i\|)$$


où $\eta_{m,1}(r) = \frac{\Gamma(1/2-m)}{2^{2m} \sqrt{\pi} (m-1)!} r^{2m-1}$ sont les composantes tortueuses

ϕ_j sont les composantes complètement lisses (par exemple $\phi_1(x)=1$, $\phi_2(x)=x$)

$\gamma_{j,k}, \delta_{i,k}$ sont les paramètres à estimer

m est la dimension de la base des splines

L'intérêt de ces splines est qu'ils ne nécessitent aucun a priori (pas de nœuds à fixer). En revanche, ils impliquent d'estimer un grand nombre de paramètres ce qui les rend très coûteux. Wood (S. N. WOOD, 2003) a donc proposé des splines de régression plaques minces qui approximent les splines plaques minces en tronquant l'espace des composantes tortueuses.

En effectuant une reparamétrisation, le problème de minimisation B.3 peut se ré-écrire sous forme matricielle comme un problème de moindre carrés pondérés et pénalisés. Il s'estime en utilisant un algorithme P-IRLS (voir section 4.3 de l'ouvrage de Wood - S. N. WOOD, 2006). Les paramètres de compromis peuvent quant à eux être estimés en minimisant l'erreur mesurée par validation croisée généralisée (voir section 4.5.3 de l'ouvrage de Wood - S. N. WOOD, 2006). Pour ce faire, la fonction `gam` de la librairie `mgcv` du logiciel  a été utilisée.

3 La régression logistique à effets mixtes (MLR)

Le modèle de régression logistique à effets mixtes permet de tenir compte d'effets individuels spécifiques. Dans le cas de l'AVC, les spécificités de chaque patient et le réglage de la mesure IRM peuvent faire fluctuer les risques associés aux paramètres IRM. Pour modéliser cela, la régression logistique à effets mixtes va introduire des paramètres spécifiques aux patients (aussi appelés effets aléatoires) :

$$\text{logit}(p_{i,id}) = \alpha + \alpha_{id} + \sum_{k=1}^K \mathcal{X}_{i,id,k}(\beta_k + b_{id,k}) \quad (\text{B.4})$$

où α_{id} et $b_{id,k}$ sont respectivement l'intercept et le paramètre de risque (associé à la $k^{\text{ième}}$ variable prédictive) spécifiques de l'individu id . Ces paramètres sont supposés suivre une loi normale de moyenne nulle et de variance τ_k . Les effets aléatoires n'étant pas observables, la vraisemblance du modèle s'écrit en intégrant la vraisemblance sur les effets aléatoires. Ainsi pour le voxel i du patient id :

$$L_v(\alpha, \beta, \tau | x_i, y_i) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} p_i^{y_i} (1 - p_i)^{1-y_i} f_{\mathcal{N}}(\alpha_{id} | 0, \tau_0) \dots f_{\mathcal{N}}(b_{id,K} | 0, \tau_K) d(\alpha_{id}) \dots d(b_{id,K})$$


L'estimation du modèle est difficile puisqu'elle nécessite une intégration numérique par effet aléatoire. Le principe général consiste à utiliser la quadrature de Gauss Hermite pour approximer chaque intégrale par une somme pondérée. On obtient alors une somme de vraisemblances pondérées et pénalisées par $f_{\mathcal{N}}(b_{id,k} | 0, \tau_k)$, la distance de l'effet aléatoire aux effets moyens. Chaque terme peut se réécrire sous la forme :

$$\|\sqrt{\text{Var}[\mathbf{p}]}^{-1}(\mathbf{y} - \mathbf{p})\|^2 + \|\mathbf{b}\|^2$$

ce qui se minimise en utilisant un algorithme P-IRLS adapté (voir le document de Bates - BATES, 2010a - chapitre 5 et la présentation associée - BATES, 2010b - pour plus de détails). Étant donné la complexité du problème d'optimisation, les algorithmes actuels ne sont pas complètement fiables lorsque la variance des effets aléatoires est grande ou que le nombre d'observations est petit (ZHANG et al., 2011).

Les modèles mixtes permettent d'estimer les effets aléatoires mais pas de les prédire pour un nouveau patient. La prédiction se fait alors sur la distribution marginale, c'est-à-dire après intégration sur les effets aléatoires :

$$p_{i,id} = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \text{logit}^{-1} \left(\alpha + \alpha_{id} + \sum_{k=1}^K \mathcal{X}_{i,id,k}(\beta_k + b_{id,k}) \right) f_{\mathcal{N}}(\alpha_{id} | 0, \tau_0) \dots f_{\mathcal{N}}(b_{id,K} | 0, \tau_K) d(\alpha_{id}) \dots d(b_{id,K})$$


En pratique nous avons utilisé la fonction `glmer` du package *lme4* du logiciel . La fonction de prédiction associée n'effectuait pas l'intégration numérique mais se contentait d'ignorer les effets aléatoires (approximation de la distribution des effets aléatoires à son mode) :

$$p_{i,id} = \text{logit}^{-1} \left(\alpha + \sum_{k=1}^K \mathcal{X}_{i,id,k} \beta_k \right)$$

4 La régression autologistique (autoLR)

La régression autologistique permet de tenir compte d'une corrélation spatiale dans le devenir des observations. Pour cela, elle introduit un terme d'autocorrélation spatial dans le prédicteur linéaire :

$$\text{logit}(p_i) = \alpha + \sum_{k=1}^K \mathcal{X}_{i,k} \beta_k + \eta W_i Y$$

où $Y = (Y_i)_{i \in \{1, \dots, n\}}$, η est le paramètre d'autocorrélation spatial réglant l'intensité du lien entre les devenirs d'observations proches et où W_i définit le voisinage du voxel i : $W_{i,j}$ est non nul si et seulement si les voxels i et j sont voisins. Les observations n'étant plus indépendantes, la vraisemblance ne peut plus se décomposer comme le produit des vraisemblances associées à chaque observation. Cela complexifie notablement l'estimation du modèle. Même si des approches Monte Carlo ont été développées, elles sont coûteuses et peu envisageables avec un nombre important de données ($n > 40\,000$ - HUGHES et al., 2011). Une version linéarisée du modèle autologistique utilisant la méthode des moments (KLIER et al., 2005, fonction `splogit` de la librairie  McSpatial) a été testée mais elle ne donnait pas des résultats satisfaisants, probablement à cause de la corrélation spatiale très élevée dans notre application. La méthode retenue a consisté à changer de fonction objectif en utilisant une pseudo vraisemblance conditionnelle (PVC). Cette approche proposée par Besag (BESAG, 1975) consiste à supposer les sites indépendants les uns des autres conditionnellement à l'espérance des valeurs voisines :

$$\mathbb{P}[Y|\mathcal{X}, \alpha, \beta, \eta] = \prod_{i=1}^n \mathbb{P}[Y_i | Y_{N(i)}, \mathcal{X}_i, \alpha, \beta, \eta]$$

où $N(i)$ désigne l'ensemble des voisins du voxel i . Le logarithme de la pseudo-vraisemblance s'écrit alors :

$$l_v^{PVC}(\alpha, \beta, \eta | Y = y, \mathcal{X} = x) = \sum_{i=1}^n \log \frac{\exp(y_i(\alpha + \beta x_i + \eta W_i y))}{1 + \exp(y_i(\alpha + \beta x_i + \eta W_i y))}$$

Cet estimateur est dans le cas Gaussien consistant mais moins efficace que le maximum de vraisemblance (GAETAN et al., 2010, section 5.4.2). L'estimation se ramène alors à une régression logistique classique. Pour la prédiction les valeurs de y voisines n'étant pas disponibles, une méthode de plug-in itérative a été utilisée (adaptée de l'article d'Augustin - AUGUSTIN et al., 1996).

L'estimation des paramètres pour la prédiction se fait suivant l'algorithme suivant :

1. ajuster une première régression logistique sans terme spatial
2. tant qu'il reste de l'auto-corrélation spatiale dans les résidus, répéter :
 - ajuster un nouveau modèle logistique en utilisant les prédictions du modèle précédent pour le prédicteur spatial
 - évaluer l'auto-corrélation des résidus

Une fois les différentes régressions logistiques estimées, la prédiction est réalisée en effectuant les prédictions dans l'ordre d'ajustement des modèles logistiques. En pratique l'algorithme convergeait très rapidement puisque deux régressions logistiques suffisaient à rendre non significative l'autocorrélation spatiale des résidus.

5 Les arbres de classification (CART)

Un arbre de classification est une méthode de classification découpant successivement l'espace formé par les prédicteurs pour définir les zones associées à chaque classe à identifier (voir figure B.1). Il existe plusieurs manières de construire un arbre de décision suivant la fonction objectif considérée ou la façon de sélectionner les variables (voir par exemple l'article de Loh - LOH, 2011). Nous nous limiterons aux arbres CART qui seront réutilisés par l'algorithme Adaboost.

L'arbre CART est construit en identifiant à chaque étape la meilleure coupure, c'est-à-dire la variable X et le seuil de cette variable τ_X définissant la meilleure partition des données ($A = \{i | \mathcal{X}_i \geq \tau_X\}$ contre $B = \{i | \mathcal{X}_i < \tau_X\}$) au sens de l'indice de Gini. Ce dernier évalue l'homogénéité des groupes A et B par la formule :

$$gini = p_A(1 - p_A) + p_B(1 - p_B)$$

avec p_A (respectivement p_B) les proportions d'observations de la partition A (resp. B) appartenant à la même classe, ici la classe des observations nécrosées (resp. saines). Une fois la coupure définie deux branches sont formées, chacune contenant les données divisées en deux selon la coupure. Si une branche est suffisamment homogène, l'algorithme s'arrête pour cette branche. Sinon une autre coupure est recherchée pour former deux branches plus homogènes. La probabilité d'appartenir à une classe est alors la proportion d'observations de cette classe dans la branche finale. Le critère d'arrêt de l'arbre inclut le nombre minimal d'observations pour diviser une branche et éventuellement le nombre maximal de coupures le long d'une branche.

Cette méthode a l'intérêt d'être numériquement efficace et très souple puisqu'elle est naturellement non linéaire et peut induire des interactions entre variables prédictives. En pratique, elle est souvent instable (un petit changement dans les données modifie notablement l'arbre) et surapprend trop. Des méthodes d'ensembles, comme l'algorithme Adaboost, ont été proposées pour améliorer le compromis biais variance des arbres de décision.

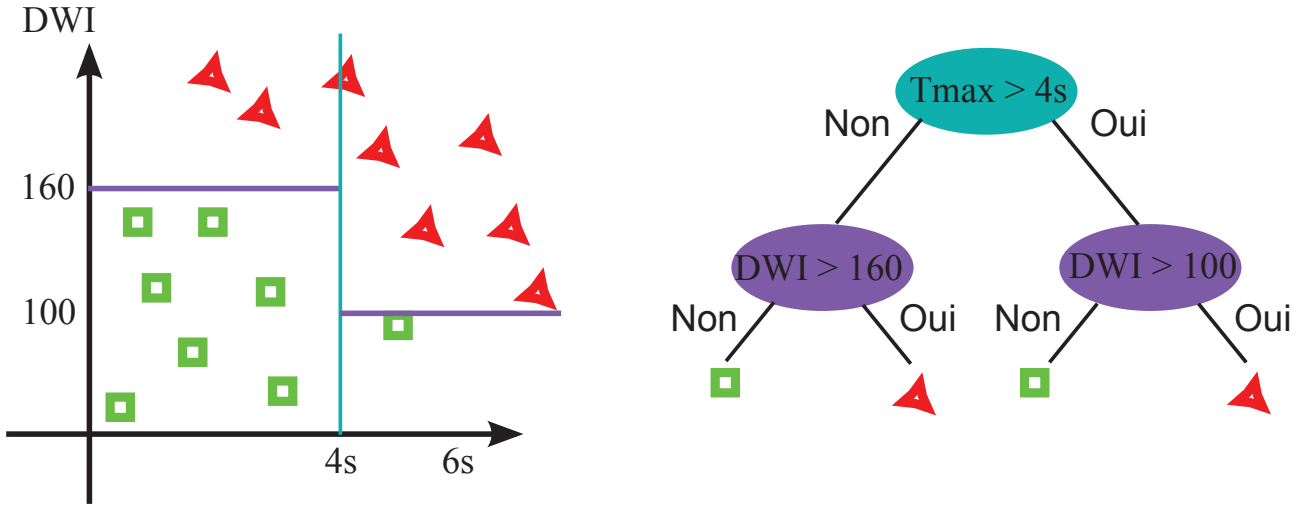


FIGURE B.1 – Exemple d'arbre de décision. À gauche espace des variables prédictives et à droite arbre de décision correspondant.

6 L'algorithme Adaboost (ADA)

L'algorithme Adaboost améliore les prédictions obtenues avec un arbre de décision en estimant des arbres supplémentaires sur les observations mal classées par les arbres précédents. La prédiction finale est alors la moyenne des prédictions de chaque arbre pondérée par la fiabilité de chaque arbre (figure B.2). Ainsi chaque nouvel arbre va apporter sa contribution à la variable de prédiction Y^* en pondérant sa propre variable de décision Y_m^* par sa fiabilité τ_m :

$$Y^*(\mathcal{X}_i) = \sum_{m=1}^{ntree} \tau_m Y_m^*(\mathcal{X}_i)$$

Plusieurs versions de l'algorithme Adaboost existent, notamment suivant les fonctions d'erreur employées. Nous retiendrons ici l'algorithme nommé "Discrete Adaboost" (FRIEDMAN et al., 2000) qui minimise une fonction d'erreur exponentielle :

$$E = \sum_{i=1}^n \exp(-Y_i Y^*(\mathcal{X}_i))$$

où l'on considère ici que les valeurs prises par Y_i sont -1 ou 1 (et non plus 0 ou 1 utilisés par ailleurs). Ainsi si Y_i et Y_i^* sont de même signe, l'erreur augmente de moins de 1 tandis que s'ils sont de signes différents, l'erreur augmente de plus de 1. Le meilleur classificateur à utiliser pour le $m^{\text{ième}}$ arbre doit alors minimiser (BISHOP, 2007, pages 659-661) :

$$E_m = \sum_{i=1}^n w_m(\mathcal{X}_i) \mathbb{1}_{Y_m^*(\mathcal{X}_i) Y_i < 0}$$

où $\mathbb{1}_{Y_m^*(\mathcal{X}_i) Y_i < 0}$ indique si la classe prédite est fautive et les poids, initialisés à 1 pour le premier

arbre, doivent respecter la relation de récurrence :

$$w_m(\mathcal{X}_i) = \begin{cases} w_{m-1}(\mathcal{X}_i) & \text{si } Y_m^*(\mathcal{X}_i)Y_i \geq 0 \\ w_{m-1}(\mathcal{X}_i)e^{\tau_m} & \text{sinon} \end{cases}$$


La fiabilité de l'arbre est alors :

$$\tau_m = \log \left(\frac{1 - \epsilon_m}{\epsilon_m} \right) \quad \text{où } \epsilon_m = \frac{\sum_{i=1}^n w_m(\mathcal{X}_i) \mathbb{1}_{Y_m^*(\mathcal{X}_i)Y_i < 0}}{\sum_{i=1}^n w_m(\mathcal{X}_i)}$$

L'algorithme Adaboost est très proche de l'estimation d'une régression logistique additive estimée avec une méthode de descente de gradient (FRIEDMAN et al., 2000). En effet, à l'ajout de chaque arbre on minimise une version approchée de la fonction objectif à l'aide de moindres carrés pondérés. Par ailleurs, le minimum de la fonction de perte est :

$$Y^* = \frac{1}{2} \log \frac{\mathbb{P}[Y = 1|\mathcal{X}]}{\mathbb{P}[Y = -1|\mathcal{X}]}$$

ce qui correspond, à un facteur 2 près, à la probabilité estimée par la régression logistique.

L'ajustement de l'algorithme Adaboost a été réalisé avec la fonction `ada` de la librairie `ada` du logiciel . L'algorithme Adaboost permet certes d'introduire une non-linéarité et des interactions entre les variables prédictives mais il suppose l'indépendance des observations ainsi qu'un choix correct du nombre d'arbres (*ntree*).

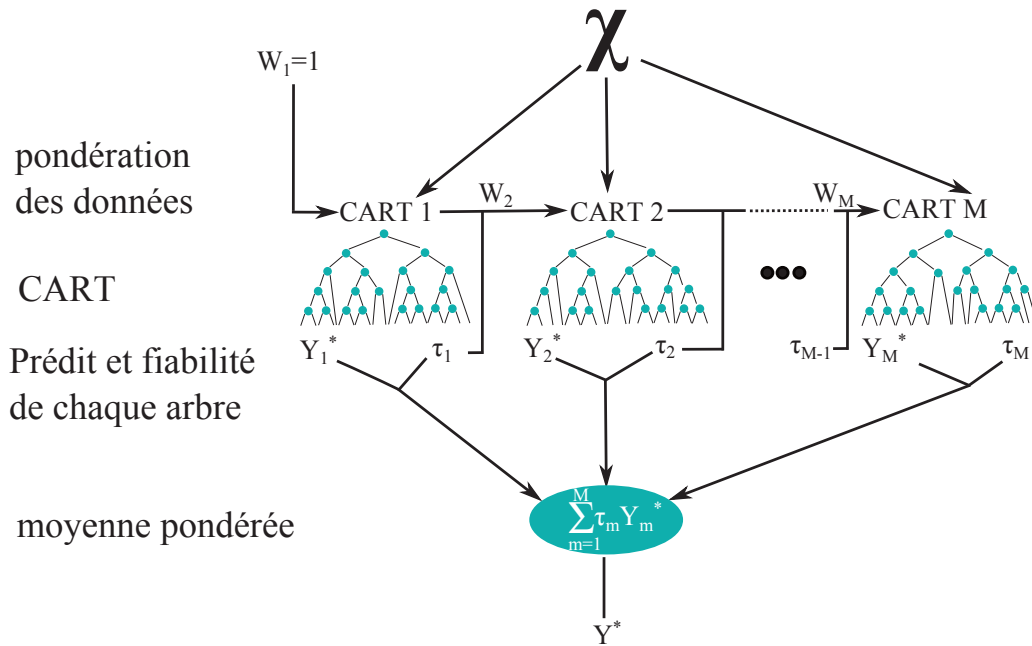


FIGURE B.2 – Schéma de fonctionnement de l'algorithme Adaboost. τ est un indicateur de la fiabilité globale de chaque arbre et est calculé en fonction de l'erreur de classification de l'arbre sur les données pondérées.

Annexe C: Correction des probabilités à posteriori du groupe lésion par un algorithme de croissance de régions

La présence de structures très intenses en T2 FLAIR ne correspondant pas à l'AVC (voir par exemple figure III.1 page 95, quatrième ligne) va générer des artéfacts de segmentation. Ces artéfacts sont généralement spatialement éloignés de la lésion et peuvent donc être éliminés avec un critère spatial. Pour cela un voxel de chaque groupe spatial de lésion est préalablement indiqué à l'algorithme de segmentation. À la fin de chaque étape E de l'algorithme de segmentation, les probabilités d'appartenance a posteriori sont corrigées :

- en considérant la probabilité d'appartenance au groupe lésion comme variable d'intensité, un algorithme de croissance de régions est initialisé avec les voxels indiqués pour chaque groupe spatial.

exemple : la lésion du patient "009" est composée de 5 groupes spatiaux. L'algorithme a donc été initialisé à l'aide de 5 voxels.

- la probabilité d'appartenance a posteriori à la lésion est filtrée en utilisant un noyau gaussien au regard de la distance de chaque voxel à la zone de lésion obtenue par l'algorithme de croissance de régions. Les probabilités a posteriori des voxels à l'intérieur de la zone est inchangé tandis que les probabilités a posteriori des voxels à l'extérieur de la zone est diminué.

exemple : un voxel situé à 10 mm de la zone de lésion définie par l'algorithme de croissance de régions verra sa probabilité a posteriori d'appartenir à la lésion divisée par 16.

- les probabilités a posteriori sont renormalisées afin que, pour chaque voxel, la somme des probabilités d'appartenance aux différents groupes vaille 1.

exemple : si le voxel précédent avait des probabilités d'appartenance aux groupes de (0,01 ; 0,05 ; 0,19 ; 0,75) puis (0,01 ; 0,05 ; 0,19 ; 0,047) après filtrage, la normalisation conduira au vecteur de probabilités (0,038 ; 0,19 ; 0,72 ; 0,047).

Les deux premières étapes sont illustrées sur la figure C.1. Il existe différentes versions de l'algorithme de croissance de régions dans la littérature. Nous avons retenu la version de (REVOL et al., 1997) pour deux raisons :

- l'utilisation d'un critère d'homogénéité de la zone fondé sur la variance des observations plutôt que sur une valeur minimum nous paraissait plus satisfaisant.

- cette méthode offre la possibilité d'exclure des voxels inclus à tort lors d'une itération précédente, ce qui la rend moins sensible à l'initialisation.

Cette approche nécessite de fixer un seul paramètre : la variance des probabilités à posteriori dans la zone de lésion. Des critères ont certes été proposés pour estimer ce paramètre (REVOL-MULLER et al., 2002) mais sur notre application ils ne fournissaient pas des résultats satisfaisants. Nous avons donc choisi d'initialiser la variance à 0,1 puis de mettre à jour sa valeur à chaque itération de l'algorithme EM, en utilisant la variance des probabilités à posteriori non corrigées.

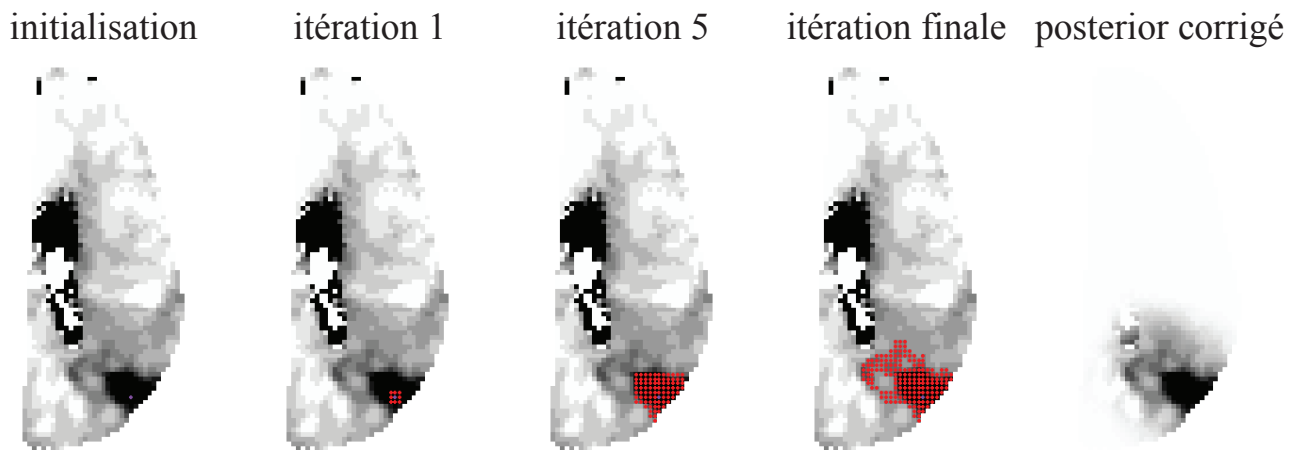


FIGURE C.1 – Étapes de correction des probabilités a posteriori d'appartenance au groupe de lésion sur un exemple fictif. Les quatre premières coupes illustrent plusieurs itérations de l'algorithme de croissance de régions. La dernière coupe montre le résultat du filtrage des probabilités à posteriori.

RÉSUMÉ :

L'Accident Vasculaire Cérébral (AVC) est une maladie grave pour laquelle des critères très stricts encadrent l'administration du traitement curatif en phase aiguë. Ces critères limitent drastiquement l'accès à ce traitement : on estime que seuls 10% des patients atteints d'un AVC en bénéficient. L'objectif de ce travail est de proposer un modèle prédictif de l'évolution de l'AVC qui permette d'identifier le volume de tissu à risque de chaque patient. Ce volume, qui correspond au bénéfice potentiel du traitement, permettra de mieux orienter le médecin dans sa décision de traiter. Pour répondre à cet objectif nous nous intéressons aux problématiques d'évaluation de modèles prédictifs dans un contexte de faible prévalence, de modélisation prédictive sur données spatiales, de prédiction volumique en fonction de l'option de traitement et de segmentation d'images en présence d'artéfacts. Les outils développés ont été rassemblés au sein d'une librairie de fonctions du logiciel R nommée MRIaggr.

MOTS-CLÉS :

Évaluation de biomarqueurs, modèles prédictifs, segmentation d'images, imagerie par résonance magnétique, accident vasculaire cérébral, logiciel R.

ABSTRACT :

Stroke is a serious disease that needs emergency health care. Due to potential side effects, the patients must fulfil very restrictive criteria for eligibility to the curative treatment. These criteria limit drastically the accessibility to treatment : currently, an estimated 10% of stroke patients are treated. The purpose of this work was to develop a statistical framework for stroke predictive models. We deal with assessing predictive models in a low-prevalence context, building predictive models for spatial data, making volumic predictions depending on the treatment option, and performing image segmentation in presence of image artefacts. Tools developed in this thesis have been collected in an R package named MRIaggr.

KEYWORDS :

biomarker assessment, predictive modelling, image segmentation, magnetic resonance imaging, stroke, R software.

DISCIPLINE/FIELD :

Biostatistique / Biostatistics

ADRESSE DU LABORATOIRE / LABORATORY ADDRESS :

Laboratoire Biostatistique Santé (UMR CNRS 5558 UCBL)
Centre Hospitalier Lyon Sud, Bâtiment 4D
69495 Pierre Benite Cedex France